AWARD NUMBER: W81XWH-10-1-1006

TITLE: Molecular Profiles for Lung Cancer Pathogenesis and Detection in U.S. Veterans

PRINCIPAL INVESTIGATOR: Steven M. Dubinett, M.D.

CONTRACTING ORGANIZATION: University of California, Los Angeles,
Los Angeles, CA 90095

REPORT DATE: December 2014

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
December 2014	Final	20 Sept 2010 – 19 Sept 2014

4. TITLE AND SUBTITLE

Molecular Profiles for Lung Cancer Pathogensis and Detection in U.S. Veterans

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-10-1-1006

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Steven M. Dubinett, M.D.
Pierre Massion, M.D.
Ignacio Wistuba, M.D.
Avrum Spira, M.D.
E-Mail: sdubinett@mednet.ucla.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of California, Los Angeles

Los Angeles, CA 90095-0001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Uncertainty about the clinical behavior of a premalignant lesion can lead to either inappropriate inaction or inappropriate aggressive treatment, either of which can result in harm to the patient. The main goal of this DOD research program was to shed light on the molecular mechanisms of airway "field cancerization" and development of premalignant lesions. During this grant, we investigated and began to understand complex mechanisms that lead to field of cancerization, including genomic aberrations and microRNA and gene expression changes. We have characterized the transcriptomic architecture of the airway field cancerization in early-stage NSCLC and validated the expression of novel field cancerization markers in airways and tumors. Furthermore, we understood how the molecular field of injury evolves spatiotemporally and identified gradient profiles in the field cancerization that highly embody the nearby lung tumors. We identified field of cancerization biomarkers that are specific to NSCLC vs benign pulmonary nodules and can detect lung cancer when assessed in minimally invasive sites in the airways. We anticipate that our results will allow for the development of chemopreventive and/or therapeutic strategies that are effective during the window of opportunity when lung cancer is still subclinical. Future translation of these results could lead to a paradigm shift in the way in which we address the clinical problem of the individual at high risk for lung cancer.

15. SUBJECT TERMS
Field cancerization; lung carcinogenesis; gene expression profiles; mutational analysis; premalignancy; progression; lung cancer stem cells

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	77	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

1. INTRODUCTION

Lung cancer is the leading cause of death from cancer in the US and the world, accounting for 28% of all cancer deaths in men and women [2]. Lung-cancer associated mortality has remained essentially unchanged over the last 3 decades, in part because the majority of lung cancers present at an advanced stage. Since 1972, when Congress declared war on cancer, lung cancer's 5 year survival rate has remain unchanged at 15% while the 5 year survival rates for breast, prostate and colon cancers have risen to 88%, 99% and 65% respectively [2]. The major cause of lung cancer is smoking, yet only 10-15% of smokers develop lung cancer [3]. Tobacco addiction and exposure to other lung cancer carcinogens are serious problems among military personnel and war veterans. In 2005, the DOD reported that over 32% of military personnel smoke. Although this is a substantial decrease when compared to a survey done in 1980, the smoking rate among the general population in 2005 was only 21% according to a report by the Centers for Disease Control and Prevention. Smoking among military personnel has increased since the late 1990s in association with conflicts in Afghanistan and Iraq and is 50% higher in deployed vs. non-deployed personnel with smoking rates in 20-25 year olds of 37% vs. 20% in the civilian population. Even conservative estimates place the cost of lung cancer to the military at $1 billion a year and this cost will only increase with this wave of new smokers. The optimal treatment for non-small-cell lung cancer (NSCLC) is surgical resection; however, 75% of patients are ineligible because of advanced disease.

The results of National Lung Screening Trial (NLST) that utilized low dose helical computed tomography (LDCT) screening in a population of heavy current or former smokers were recently published and demonstrated a 20% reduction in lung cancer mortality and a 6.7% decrease in all-cause mortality with LDCT relative to CXR [4, 5]. This is the single most significant advance in reducing lung cancer mortality since 1964, when the United States Surgeon General first publicized the causal relationship between lung cancer and cigarette smoking, prompting a decline in active smoking in the US. However, with the reported 96.4% false positives rate there is a pressing need to develop reliable molecular biomarkers to supplement the radiologic imaging to improve the sensitivity of early detection of lung cancer. Based on the notions that: 1) smoking-induced injury alters mRNA and microRNA (miRNA) expression profiles in airway epithelium [1, 6, 7], and 2) these changes can be detected and serve as biomarker for early detection of lung cancer [8, 9], in the current project we have focused on investigation of the molecular events associated with field cancerization to identify biomarkers of early lung carcinogenesis.

In Specific Aim 1 of this program, high-throughput microarray mRNA expression analyses have been performed on cytological specimens (brushings) obtained at intraoperative bronchoscopy from the main carina and main ipsilateral bronchus, as well as on specimens obtained at lobectomy procedures from the main lobe bronchus (adjacent to SCCs), sub-segmental bronchus (adjacent to adenocarcinomas) and from the resected NSCLC tumors. Towards this aim, we compared and contrasted global gene expression patterns across all the specimens from the entire field and corresponding NSCLC tumors. We developed lung adenocarcinoma and SCC field cancerization signatures signifying the differential mRNA expression patterns between the carina and the subsegmental bronchus and main lobe bronchus, respectively. In addition, similar expression profiles between the carina and resected NSCLC tumors have been integrated with available gene expression data of bronchial brushings from the main carina isolated at various time points post-surgery from 40 NSCLC patients; Department of Defense (DoD) VITAL patients. Promising markers derived from this study are being validated at the mRNA and protein level in histological tissue specimens. Moreover, we performed RNA-sequencing and microarray profiling of nasal epithelia, airway epithelial cells collected from both bronchoscopy and lobectomy specimens as well as of corresponding tumors (NSCLC patients) or benign lesions (cancer-free individuals).

In Specific Aim 2, we utilized laser capture microdissection to obtain specific cell populations (basal cells or type II alveolar cells, depending on the NSCLC histology/location) as well as premalignant lesions and epithelial components of the tumors. These cell populations have been profiled with RNA-seq to determine their gene expression signatures to increase our understanding of premalignancy. We identified the gene expression profiles that are associated with progression from a benign cell population to premalignancy and with progression from a benign cell population to true malignancy. Additionally we performed whole exome DNA sequencing of matched sets of premalignant lesions, adenocarcinoma *in situ*, lung cancer and adjacent normal tissues isolated from the same patients to determine the role of somatic mutations in lung cancer initiation and progression.

In Specific Aim 3, we utilized expression signatures and biomarkers derived from the results of aims 1 and 2 to develop and test airway-based biomarkers capable of diagnosing lung cancer in current or former smokers using minimally invasive sites.

2. KEYWORDS

Lung cancer, biomarkers, miRNA profiling, protein profiling, expression analysis, whole exome sequencing, early detection.

3. OVERALL PROJECT SUMMARY

Specific Aim 1: To increase our understanding of the molecular basis of the pathogenesis of lung cancer in the "field cancerization" that develops in current and former smokers.

Summary of Research Findings

A. To increase our understanding of the molecular basis of the pathogenesis of lung cancer in the "field cancerization" that develops in current and former smokers.

Collection of airway epithelial samples from both bronchoscopy and lobectomy specimens from smokers with and without lung cancer.
Over the course of this grant, we were able to recruit 37 subjects undergoing resection of lung tumor or benign lung lesions across four institutions (MDAnderson, Boston University, University of California Los Angeles, and Vanderbilt University) for the purposes of studies in Aim 1 (**Table 1**). We collected biospecimens from seven locations: nasal epithelium, proximal and distal bronchial airway epithelium obtained at bronchoscopy (ipsilateral and contralateral to the tumor) as well as the tumor/benign lesion, adjacent normal parenchyma, and sub-segmental bronchial epithelium at time of lobectomy (**Figure 1**).

	Cancer (n =28)	No Cancer (n=9)	p value
Age	65 (11)	52 (16)	0.05
Gender	10 M, 18 F	2 M, 7 F	0.69
Smoking Status	5 C, 21 Fr, 2 N	2 C, 4 Fr, 3 N	0.11
Cancer type	16 ADC, 12 SCC	N/A	N/A

Table 1. Subject Demographics. Age is listed as the average (standard deviation). C, current; Fr, former; N, never. ADC, adenocarcinoma; SCC, squamous cell carcinoma.

High throughput gene expression profiling (Aim1A).
Total RNA was isolated and profiled by microarray (Affymetrix GeneChip Human Gene 2.0 ST). Microarrays were normalized and outliers were removed based on the Relative Log Expression (RLE), Normalized Unscaled Standard Error (NUSE) quality metrics and Principal Component Analysis (PCA). Using linear modeling we compared patients with cancer to those with benign lung disease to identify genes with cancer-associated gene-expression across sites. We identified two patterns: 1) cancer-associated expression differences that changed with distance from tumors (field of cancerization, 47 genes), and 2) genes consistently associated with cancer across the sites (field of injury, 106 genes). **Figure 2** shows the field cancerization effect in which cancer-associated changes in gene expression in the proximal airway fall off as distance from the tumor increases. This pattern is not observed in patients without cancer. The second pattern of gene expression (genes consistently changed across the thoracic airway depending on the presence or absence of a tumor) are significantly enriched by GSEA among genes we previously found to be altered in the large airway of patients with cancer (pval < 0.001, **Figure 3**). In contrast, genes that change in a gradient like manner (field of cancerization, 47 genes) were not enriched in that previously published large airway dataset. These findings suggest that the genes we identified changing consistent throughout the airway (field of injury, 106 genes) hold the potential to serve as early detection biomarkers for lung cancer in the relatively accessible proximal airway.

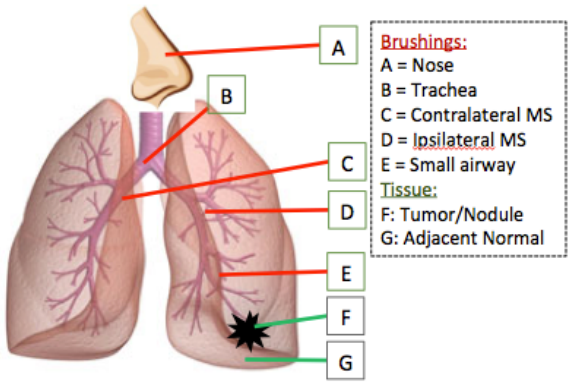


Figure 1. Overview of biospecimen collection sites.

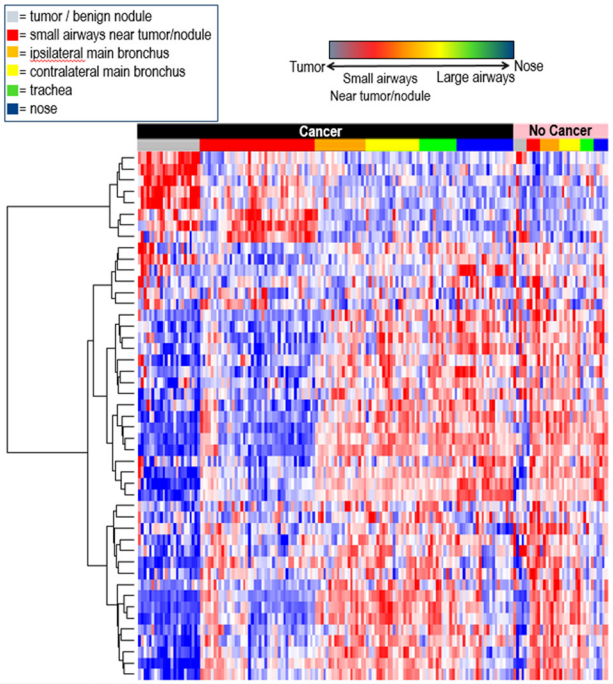


Figure 2. Gene expression changes associated with the "Field of Cancerization Effect". 47 genes shown are p < 0.005 for the interaction term and p < 0.005 for the cancer term.

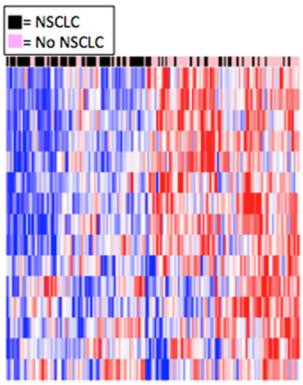


Figure 3. Genes consistently changing throughout the field of injury are able to distinguish large airway epithelium from smokers with and without lung cancer in an independent dataset [1]. Heatmap showing expression levels for leading edge genes from our "field of injury" signature in bronchial airway samples collected from smokers undergoing bronchoscopy for suspicion of lung cancer. Black = patients diagnosed with cancer, Pink = patients diagnosed with benign lung disease. Red = increased expression, Blue=decreased expression.

Additionally, we have performed RNA-seq on a subset of the specimens from subjects (n = 17) run on arrays in order to both confirm the gene expression signatures identified by microarrays as well as identify additional transcripts within the field of injury not interrogated by the microarrays.

B. Gene expression analysis of the field cancerization in early-stage NSCLC patients

Spatiotemporal field of cancerization expression profiles.

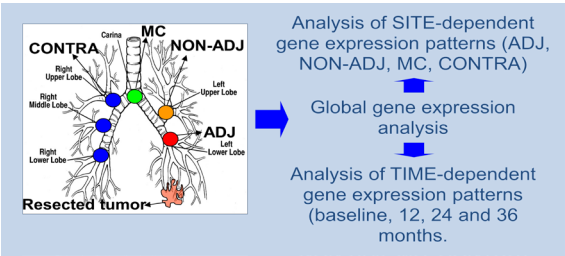


Figure 4. Schematic illustrating sampling and collection of bronchial brushes from different airway sites at different time points following resective surgery from early stage smoker NSCLC patients enrolled in the phase II cancer surveillance Vanguard study.

Gene expression alterations in response to cigarette smoke have been characterized in normal-appearing bronchial epithelium of healthy smokers, and it has been suggested that adjacent histologically normal tissue displays tumor-associated molecular abnormalities [10, 11]. We sought to delineate how the field of injury evolves in space and in time in NSCLC patients following curative surgery. In **Years 01-02** of this program, we performed expression profiling of samples from smoker patients who were accrued into a surveillance clinical trial for annual follow-up and bronchoscopies within 1 year after definitive surgery. Bronchial brushings and biopsies were obtained from six different sites in the lung at the time of inclusion in the study and at 12, 24, and 36 months after the first time point (**Figure 4**) [12]. Our study pinpointed spatial and temporal cancer-associated expression alterations in the molecular field of injury of patients with early-stage NSCLCs after definitive surgery that enrich the molecular definition of the airway field of cancerization and establish new paradigms for the patient at risk for lung cancer [12, 13] (see **Reportable Outcomes** and **Appendix**).

Transcriptomic architecture of the adjacent airway field of cancerization.

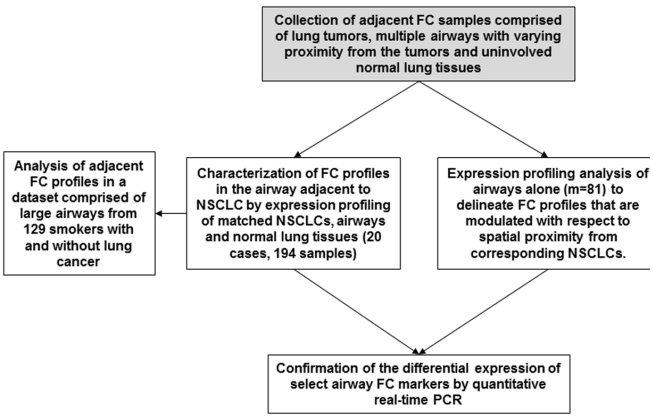


Figure 5. Schematic of the transcriptomic analysis of the airway field cancerization in NSCLC.

We also sought to characterize the yet unknown global molecular and adjacent airway field cancerization in early-stage NSCLC. In **Years 01-03** of this program, we performed whole-transcriptome expression profiling of resected early-stage (I-IIIA) NSCLC specimens (n=20) with matched tumors, multiple cytologically controlled normal airways with varying distances from tumors and uninvolved normal lung tissues (n=194 samples) using the Affymetrix Human Gene 1.0 ST platform. A schematic of the recently published [14] study's design is represented in **Figure 5**. We identified differentially expressed gene features (n = 1661) between NSCLCs and airways compared with normal lung tissues, a subset of which (n = 299), after gene set enrichment analysis, statistically significantly ($P < .001$) distinguished large airways in lung cancer patients from airways in cancer-free smokers. In addition, we identified genes (n = 422) statistically significantly and progressively differentially expressed in airways by distance from tumors that were found to be congruently modulated between NSCLCs and normal lung tissues. Furthermore, *LAPTM4B*, with statistically significantly increased expression ($P < .05$) in airways with shorter distance from tumors, was up-regulated in human immortalized cells compared with normal bronchial epithelial cells ($P < .001$) and promoted anchorage-dependent and -independent lung cancer cell growth. Our studies demonstrate that the adjacent airway field of cancerization comprises both site-independent profiles as well as gradient and localized airway expression patterns [14]. Profiling of the airway field of cancerization may provide new insights into NSCLC oncogenesis and molecular tools for detection of the disease [14] (see **Reportable Outcomes** and **Appendix**).

Our microarray analysis also pinpointed to a decreased expression of the lung-specific tumor suppressor gene G-protein coupled receptor family C, group 5, member A (*GPRC5A*) in tumors and airways compared to uninvolved normal lung tissue. *GPRC5A* was identified as a lung-specific tumor suppressor gene evidenced by spontaneous and tobacco-driven adenocarcinoma formation in mice with knockout of both of the gene's alleles [15-19]. *GPRC5A* was shown to exert its tumor suppressive function, in part, by inhibition of nuclear factor-kappa B ($NF\kappa B$) and downstream inflammation [16]. In **Year 02** of this program, we analyzed *GPRC5A* expression in the molecular field cancerization associated with chronic obstructive pulmonary disease (COPD), a risk factor for lung cancer that is typically associated with inflammation [20]. Quantitative real-time PCR (QRTPCR) in an independent set of samples demonstrated that *GPRC5A* expression was significantly decreased in the molecular localized field cancerization [20]. GPRC5A airway expression was highest in airways from cancer- and COPD-free smokers, decreased in airways of COPD patients ($p = 0.004$) and lowest in airway epithelia of COPD patients with adenocarcinoma and SCC ($P < 0.0001$) [20] (see **Reportable Outcomes** and **Appendix**).

Somatic point mutations and allelic imbalance in the airway field of cancerization.

Our findings on the transcriptomic architecture of the airway field of cancerization in this DoD-funded program prompted us to perform preliminary studies in **Year 04** to investigate somatic genomic alterations in the airway field. In a pilot study, we performed deep sequencing of a panel of 409 cancer-associated genes using the Ion Proton sequencing platform (Life Technologies), in samples from several lung adenocarcinoma (LUAD) patients consisting of multi-region tumor core needle biopsies (CNBs), normal-appearing small airways (SA) adjacent to the LUADs, normal large airways (LA), nasal epithelia and white blood cells. We achieved an average depth of 900X and 80% of the samples achieved >500X across 80% of the capture region. We applied the Genome Analysis Toolkit (GATK)-based workflow [21], running MuTect [22] to infer acquired mutations with the white blood cell-derived DNA as the paired normal sample. We obtained counts of somatic variants for each non-blood sample. We found that there were some mutations were shared between the tumor and adjacent "normal" airway samples, while others were unique (private) to these regions. There appeared greater sharing between the adjacent "normal" airway and LUAD than between the tumor and either the large airway (LA) or nasal epithelium. **Figure 6A** depicts mutations in fibroblast growth factor receptor 4 (*FGFR4*, **red**) and TATA box binding protein-associated factor (*TAF1L*; **blue**) that were found to be shared between LUAD CNBs and adjacent "normal" small airways (SA). The variant alleles for the *FGFR4* mutations were found in decreasing frequency from some LUAD CNBs to the SA, LA, nasal (and was not observed in the blood), fitting the profile of a spatial field cancerization mutation (**Figure 6A; pt. A**). We also performed a pilot genome-wide SNP array profiling study and used hapLOH, a computational tool developed by the Scheet laboratory [23-25], at MD Anderson Cancer Center, that incorporates information about germline haplotypes and allows detection of aberrant DNA particularly in settings when there exists clonal mosaicism in adjacent normal tissue, which we anticipate in the field cancerization. Using hapLOH we identified somatic allelic imbalance events in the airway field cancerization. **Figure 6B** displays an event in chromosomal region 9q that is shared between the tumor and the adjacent normal-appearing airway in two LUAD patients. This event was not detected in white blood cells (**Figure 6**) or in uninvolved and distant normal lung tissues. Our findings pinpoint somatic genomic aberrations in the normal-appearing airway field of cancerization. These findings were submitted as an abstract for presentation in the forthcoming American Association for Cancer Research Annual meeting.

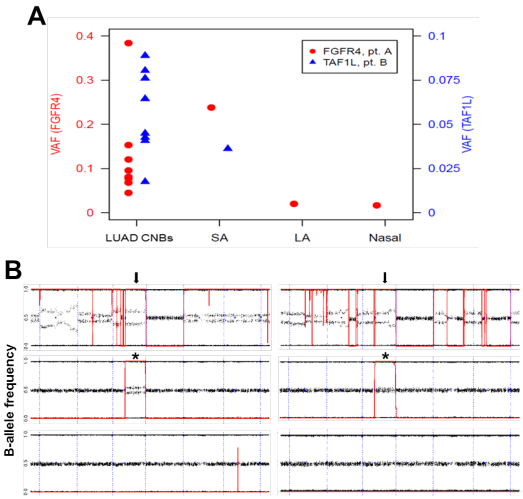


Figure 6. Genomic alterations in the airway field cancerization. A. Somatic point mutations identified by deep sequencing and present in both NSCLCs and adjacent normal airways. **B.** Identification by hapLOH (red vertical lines) of tumor-associated somatic chromosomal aberrations in "normal" airways of two patients (left and right). Black arrows and asterisks indicate allelic imbalance in 9q found in both NSCLCs (top) and airways (middle) and not found in blood (bottom). CNBs, core needle biopsies; SA, small adjacent airway; LA, large airway.

Role of LAPTM4B field cancerization marker in lung cancer pathogenesis.

One of the top aberrant gradient and localized field cancerization markers from our recent study [14] was the lysosomal membrane associated putative oncogene *LAPTM4B*. *LAPTM4B* was recently shown to promote autophagy for breast cancer cell survival and to mediate breast cancer chemoresistance [26, 27]. In **Years 03-04** of this program, we sought to study the role of *LAPTM4B* in lung cancer

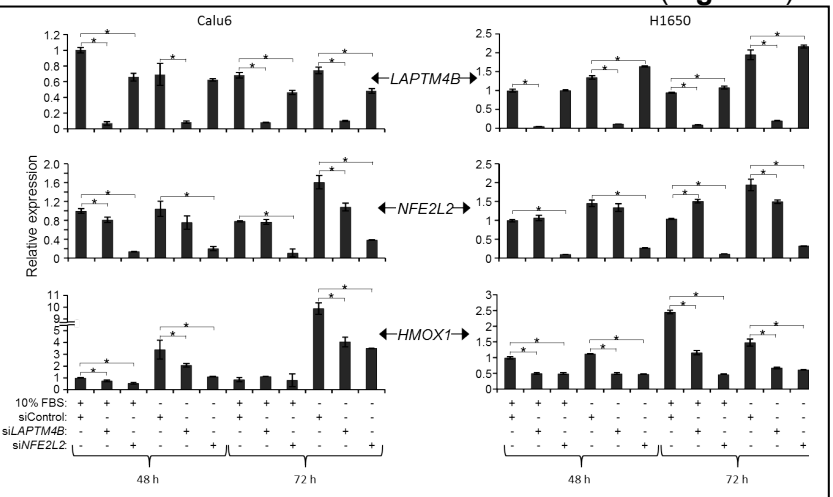


Figure 7. *LAPTM4B* promotes serum starvation-mediated expression of *NFE2L2* and heme oxygenase 1 (*HMOX1*) in lung cancer cells. QRTPCR analysis of *LAPTM4B* (top), *NFE2L2* (middle) and *HMOX1* (bottom) expression in Calu6 (left) and H1650 (right) lung cancer cells transfected with control siRNA as well as *LAPTM4B*-specific and *NFE2L2*-specific siRNA in the presence and absence of 10% FBS.

pathogenesis. As detailed in our previous annual report (**Year 03**), we already demonstrated that 1) knockdown of *LAPTM4B* suppressed anchorage-dependent and –independent lung cancer cell growth; 2) *LAPTM4B* expression, when analyzed by *in situ* hybridization in NSCLC tissues, was predictive of poor survival and prognosis; 3) the anti-growth effects of *LAPTM4B* knockdown were substantially larger when cells were cultured in medium lacking serum compared to when cells were cultured in serum-containing medium; 4) *LAPTM4B* activates cellular autophagy to promote lung cancer cell survival following serum starvation and 5) *LAPTM4B* activated and up-regulated the nuclear factor erythroid 2-like 2 (NFE2L2 also known as NRF2) transcription factor along with the NRF2-mediated response as evidenced by functional pathways analysis of expression profiles of cells with and without knockdown of *LAPTM4B*.

In **Year 04** of the program, we further explored the interplay between LAPTM4B and the NRF2-mediated stress response. QRTPCR analysis of the NRF2 target heme oxygenase 1 (*HMOX1*) demonstrated that mRNA levels of the NRF2 target *HMOX1* were significantly increased after 48 h and 72 h of serum withdrawal. Knockdown of *LAPTM4B* significantly attenuated *HMOX1* induction by serum withdrawal (**Figure 7**). In addition, knockdown of *LAPTM4B* concomitantly reduced the expression levels of the NRF2 transcription factor itself.

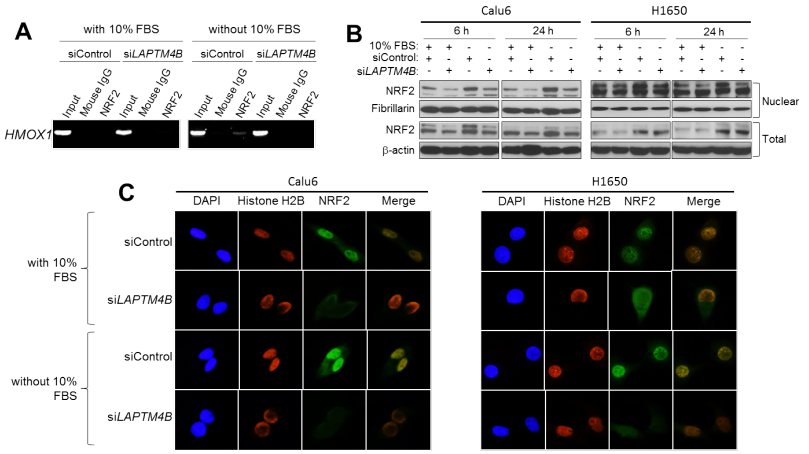


Figure 8. LAPTM4B promotes nuclear localization of the NRF2 transcription factor. A. ChIP analysis of NRF2 protein binding to HMOX1 promoter in Calu6 lung cancer cells. **B.** Western blot analysis of nuclear and total NRF2 protein following *LAPTM4B* knockdown in presence and absence of 10% FBS. **C.** ICC analysis of NRF2 nuclear localization following *LAPTM4B* knockdown in presence and absence of 10% FBS.

We then sought to assess the effect of *LAPTM4B* expression on the NRF2 transcription factor itself. We first performed chromatin immunoprecipitation (ChIP) analysis of the effect of *LAPTM4B* knockdown on the transactivation of *HMOX1* by the NRF2 transcription factor. Using primers covering NRF2-binding sites in the HMOX1 promoter, ChIP demonstrated that *LAPTM4B* knockdown decreased serum starvation-induced transactivation of *HMOX1* by NRF2 evidenced by reduced binding of the transcription factor to the *HMOX1* promoter (**Figure 8A**). Moreover and in accordance with QRTPCR and ChIP analyses, knockdown of *LAPTM4B* attenuated nuclear accumulation of NRF2 protein by serum starvation as evidenced by western blotting of nuclear cell fractions (**Figure 8B**) and by ICC analysis of the co-localization of NRF2 with the nuclear marker histone 2B (**Figure 8C**).

Furthermore, reciprocal effects were observed in cells transfected with *LAPTM4B*-overexpressing vectors. Over-expression of *LAPTM4B* significantly augmented *HMOX1* and NRF2 induction by serum starvation (**Figures 9A-C**) as well as the nuclear accumulation of NRF2 (**Figure 9D**). It is noteworthy that while modulation of *LAPTM4B* expression did not affect levels of *HMOX1* in cells at basal conditions and cultured in FBS-containing medium (**Figure 9C**), *LAPTM4B* expression positively

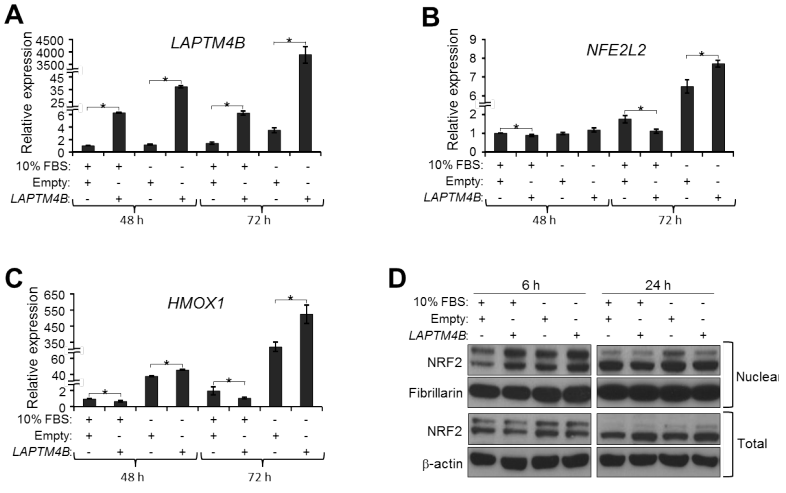


Figure 9. Over-expression of *LAPTM4B* increases NRF2 activation and *HMOX1* expression. QRTPCR analysis of *LAPTM4B* (**A**), *NFE2L2* (**B**) and *HMOX1* (**C**) mRNA levels in Calu6 lung cancer cells following *LAPTM4B* over-expression in the presence and absence of 10% FBS. **D.** Western blot analysis of nuclear and total NRF2 protein in Calu6 lung cancer cells following *LAPTM4B* over-expression in the presence and absence of 10% FBS.

controlled nuclear levels of the transcription factor NRF2 (**Figure 9D**). It is reasonable to surmise that *HMOX1* induction by NRF2 is independent on the levels of the latter transcription factor but rather dependent on serum deprivation-induced stress. Our findings point to a novel intracellular mechanism, which involves the *LAPTM4B* field cancerization marker, for control of the NRF2 transcription factor during basal conditions and cellular stress (e.g. nutrient deprivation). These findings were presented in the past 2014 American Association for Cancer Research Annual meeting [28] and are currently being prepared for publication.

C. Investigating molecular signature of lung cancer development in bronchial specimens by shotgun proteomics.

A major challenge in reducing mortality from lung cancer hinges on identifying at-risk population. The bronchial epithelium may serve as surrogate tissue for biomarker of risk. Our objective was to identify a signature of risk from proteomic alterations in the cytologically normal airway epithelium from individuals at risk for developing lung cancer. We have identified 2792 protein, the largest to our knowledge, in bronchial brushings specimens collected from risk stratified individuals by shotgun proteomics. Joncheere-Terpstra trend test among low, medium and high risk groups resulted significantly altered expression of 316 proteins (trend p <0.05) with 238 up and 78 down trends. Pathway enrichment analysis revealed up-regulation of carbohydrate metabolic pathways in the airways of high risk individuals. Expression of carbohydrate metabolic enzymes, lactate production and glucose consumption were increased in in vitro culture of human bronchial epithelial cells treated with cigarette smoke condensate. Up-regulation of these enzymes and other selected candidate proteins were validated by parallel reaction monitoring mass spectrometry and tissue microarray of histologically normal bronchial tissues from risk stratified individuals. We identified a signature of lung cancer risk assessment that may provide the basis of patient selection for surveillance programs and chemoprevention. Pathway enrichment analysis as well as in vitro data suggests possible metabolic reprogramming in the airway epithelium of high risk individuals.

Methods

Here we hypothesized that proteomic alterations in the histologically normal airway epithelium from risk stratified individuals allow us to derive a signature of risk of developing lung cancer.

Specimen collection procedure: Bronchial brushings were collected from patients under conscious sedation. The brushings were immediately dipped into 1.5 ml saline taken in a labeled eppendorf tube. The tube was kept on ice to minimize protease action. Care was taken to keep the brush specimen free from blood. Brushings in the saline was vigorously agitated by vortexing for about 10 seconds with highest speed. It was then spun 1500g for 10 minutes in a microcentrfuge with the brush inside the tube. Supernatant was removed carefully leaving as little saline as possible keeping the brush inside the tube. The pellet was stored in freezer at -80^{o}C temperature. Patients undergoing autofluorescence bronchoscopy for clinical suspicion of lung cancer agreed to provide bronchial biopsy specimens at predetermined normal sites (with normal fluorescence ratio). Biopsy specimens collected for research were snap frozen and stored in -80^{o}C freezer. Patients characteristics are in **Table 2**.

Table 2. Characteristics of individuals providing bronchial brushings for the training set

Characteristics		Risk levels low (n=5)	Medium (n=5)	High (n=5)
Age				
	Average+stdev	58.2+4.7	59.4+10.7	72+5.2
	Median (range)	58 (53-65)	59 (47-74)	69 (68-78)
Gender				
	Male	4	2	3
	Female	1	3	2
Smoking status				
	Never smoker	5	0	0
	Ex-smoker	0	5	4
	Current smoker	0	0	1
	Pack year, avg+stdev	n/a	21.5+24.1	62.8+25.4
	Pack year, median (range)	n/a	20 (1-60)	50 (50-108)

Data acquisition: Proteins were extracted and fractionated as described previously. Equal amounts (micrograms) of proteins from each specimen of each group were pooled, divided into three technical replicates and each replicate was fractionated by isoelectric focusing (IEF). Shotgun proteomic data from each of the fifteen IEF fractions were collected by LC MS/MS (Velos). A combination of database search tools were employed to identify peptide sequence present in a protein/protein groups. Normalized spectral counts of the digested proteins were considered as the basis of protein abundance measurement. Protein groups were identified by peptide matching, database search, filtration and parsimonious analyses using search tools

developed at Vanderbilt University Medical Center (18-21). A protein group may include isoforms, variants or in rare occasions, unrelated proteins containing identical peptide sequence. Lisa and Rob to write the detail about sample preparation for LC MS/MS analysis. IEF fractionation of peptide digests and LC MS/MS data acquisition and analysis. Peptide identification from LC MS/MS data, protein assembly and filtering.

Statistical Methods: Joncheere-Terpstra Trend analysis Ming to write (groups 1, 2 and 3 Trend analysis). Validation of selected identified proteins by parallel reactions monitoring (PRM) analysis. Overexpressed carbohydrate metabolic enzymes in biopsies collected from individuals who provided bronchial brushings for the training set by Western blotting using matched biopsies.

Results

We have identified 2792 protein groups in the bronchial brushings specimen of risk stratified individuals by LC MS/MS shotgun proteomics (**Table 3**). Volunteers/subjects were categorized as low, medium and high risk groups using the lung cancer screening tool. Among the identified proteins in each group 50 to 60% were identified in each of the three replicates. Each protein was identified by two or more peptide sequences. Number of protein groups identified separately in groups 1, 2 and 3 were 2116, 1949 and 2771 respectively. Reproducibility of protein identifications in all three replicates were 49%, 48% and 61% in groups 1, 2 and 3 respectively. To discover significantly dysregulated proteins in the bronchial epithelium of Individuals at risk for lung cancer development we performed Jonckheere-Terpstra trend test with 2526 identified proteins. Number of proteins with statistically significant (trend p <0.05) increasing and decreasing trends were 316. Normalized spectral counts were used as quantitative measure of protein abundance. Expression of 238 out of 316 significantly dysregulated proteins demonstrated increasing trend and that of 78 proteins demonstrated decreasing trend.

Table 3: Number of identified protein groups and spectral counts of each risk group

Risk Group	Filtered Spectra	Distinct Peptides	Protein Groups
All groups			2792
Low	24637	5812	2116
Medium	24530	5115	1949
High	36598	9579	2771

Pathway enrichment analysis reveals profound alterations in metabolic enzymes regulating the glycolytic pathway and the TCA cycle. To understand the biological relevance of these alterations in the bronchial epithelium proteome pathway analysis was performed using Webgestalt (WEB-based GEne SeT AnaLysis Toolkit) (22). Because of the unavailability of larger human bronchial epithelium proteome we used our total identified proteins (2792 proteins) as the reference proteome of bronchial epithelium instead of entire human proteome. KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis with 316 dysregulated proteins revealed dysregulation of metabolic pathways. Interestingly, enzymes of the glycolytic pathway, TCA cycle, pentose phosphate pathway, galactose, and glycogen metabolisms were overexpressed indicating very early events of possible metabolic reprogramming in the histologically normal

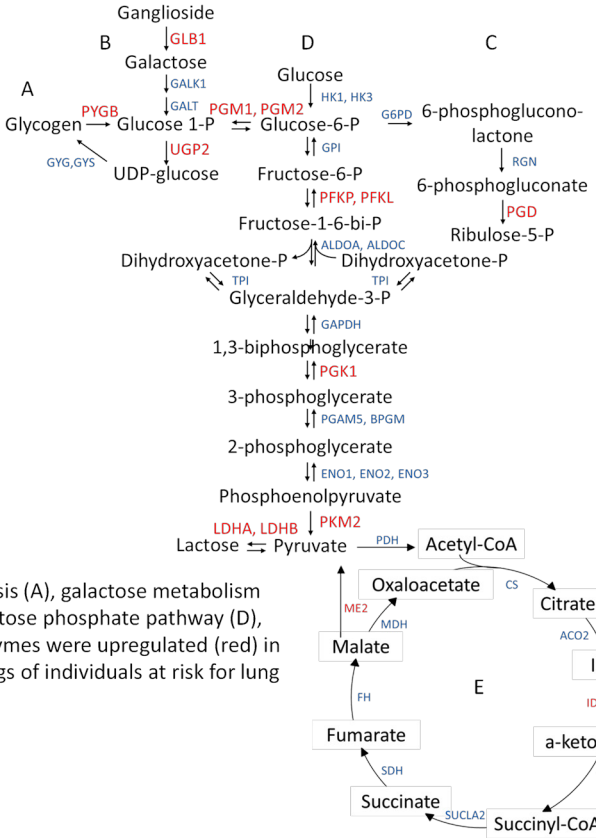


Figure 10. Glycogenolysis (A), galactose metabolism (B), glycolysis (C), pentose phosphate pathway (D), and TCA cycle (E) enzymes were upregulated (red) in the bronchial brushings of individuals at risk for lung cancer development.

bronchial epithelium of individuals at risk for lung cancer development (**Figure 10**). Moreover, six out of fourteen enzymes of the above mentioned pathways are rate limiting enzymes. Results of the Jonckheere-Terpstra trend test showing statistically significant increasing trend of the overexpressed carbohydrate metabolic enzymes among the risk groups using normalized spectral count of the replicates of each risk group are show in **Figure 11**.

Overexpression of ME2 may indicate diversion of excess TCA cycle intermediate malate to fatty acid biosynthetic pathway via pyruvate and acetyl coA. Among the 316 significantly altered proteins there were 49 metabolic enzymes. As expected enzymes of lipid metabolic pathways, fatty acid synthase (FASN), Acyl-CoA synthetase family member 3 (ACSF3) and arachidonate 15-lipoxygenase were also found to be overexpressed. Overexpression of alcohol and aldehyde metabolic enzymes ALDH1A1, ALDH3A1 and AKR1B10 in lung tumor tissues was reported previously by our group and other investigators and now we found these enzymes to be overexpressed in the histologically normal bronchial epithelium of individuals at risk for lung cancer development.

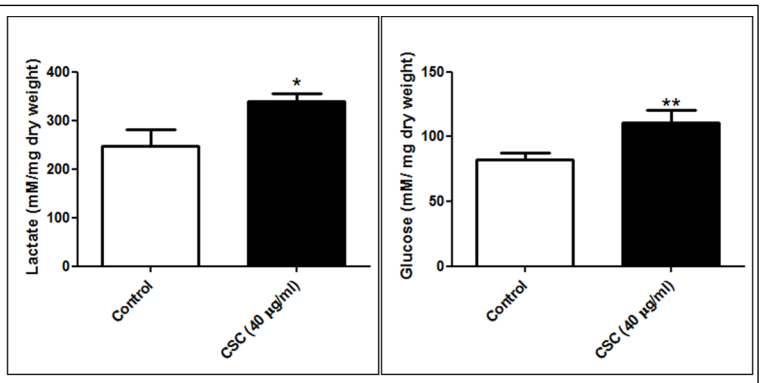


Figure 11. Increased lactate production and glucose consumption by human bronchial epithelial cell line, BEAS2B after treatment with CSC for 24 hours *in vitro*.

Pathway analysis revealed metabolism as the top pathway to be dysregulated in individuals at risk for lung cancer development. The identification of ACSF3, ALOX15 and FASN in the bronchial epithelium of at risk individuals raises the possibility of involvement of lipid metabolism in early events of lung tumorigenesis. Previously we reported overexpression of acyl-coA binding protein/Diazepam binding inhibitor (ACBP/DBI) in lung cancer tissues indicating perturbation of lipid metabolism in lung cancer. To our knowledge, overexpression of the proteins identified and validated in the current study in the bronchial epithelium of individuals at high risk for developing lung cancer was not reported before.

Validation of over expression of carbohydrate metabolic enzymes and selected candidate proteins by PRM and immunohistochemistry in bronchial the epthelium. These experiments are in progress.

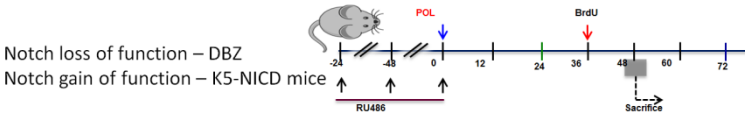
Warburg effect in the human bronchial epithelial cell lines by cigarette smoke condensate (CSC).
Given the overexpression of glycolytic enzymes including LDH we were curious to know if the level of lactate were altered in the airway epithelium of high risk individuals. As an alternative risk model we treated normal bronchial epithelial cell lines, 16HBE and BEAS2B with CSC in vitro and measured lactate and glucose concentrations in culture supernatant after 24 hours. Significantly enhanced lactate production and glucose consumption in vitro as a result of treatment of the cells with CSC (**Figure 11**) suggests a possible role of carbohydrate metabolic enzymes in lung cancer development. Whether these metabolic changes have any causative effect on the transformation of normal bronchial epithelial cells to dysplastic and finally malignant cells as Otto Warburg postulated more than half century ago remains to be elucidated.

Specific Aim 2: Evaluate the role of airway epithelium tumor-initiating stem/progenitor cells in current and former smokers.

Summary of Research Findings:

A. Assessment of the molecular profiles of tumor-initiating stem/progenitor cells from normal airway epithelium, premalignant lesions and cancer.
During the fourth year of funding, the work performed over the last 3 years and described in the previous progress report was published [29]. Briefly, we published an analysis of the transcriptomic changes that were most relevant for progression of premalignant lesions to squamous lung cancer. We next explored one of the pathways that we found to be altered with initiation and progression of squamous cell carcinoma (SCC) within individual patients. This pathway is the reactive oxygen species pathway. Cigarette smoke contains high levels

of reactive oxygen species, and this is therefore directly relevant to lung injury and repair in current and former smokers.



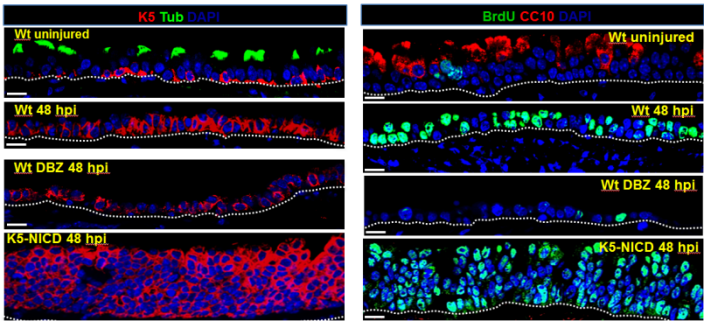


Figure 12. Role of activated Notch1 in the development of premalignant lesions of the large airways. Repair of the airway epithelium was assessed after polidocanol injury. For Notch1 activation, K5-NICD mice received intratracheal RU486 to turn on the NICD transgene in K5-expressing basal cells. For Notch1 inhibition the gamma secretase inhibitor, DBZ, was administered. At 48 hours post injury the DBZ treated mice had delayed repair of the airway epithelium with reduced proliferation, while the K5-NICD mice demonstrated excessive proliferation of basal cells with loss of differentiation resulting in the hyperplasia and dysplasia of premalignant lesions.

B. Role of reactive oxygen species in lung carcinogenesis.

We found that reactive oxygen species (ROS) in airway basal stem cells are usually maintained at low levels by antioxidants in the cell, but after injury there is flux of ROS from low levels to a higher level ("sweet spot") and it is this flux that allows the stem cell to self-renew for repair. The higher ROS levels are rapidly reduced in the stem cells to achieve homeostasis and prevent the basal stem cells from proliferating too much and producing hyperplasia and dysplasia in the form of a premalignant lesion. We further showed that this increase in ROS in basal stem cells in the injured airway results in an increase in Nrf2 expression which acts through Notch1 to promote the proliferation of airway basal stem cells. Of note, the premalignant lesions that occur when the ROS/Nrf2/Notch1 pathway is perturbed are able to spontaneously resolve, just as most premalignant lesions resolve in COPD patients. Thus, it is likely that additional genomic events are responsible for persistence of rare premalignant lesions that can then progress to invasive lung cancer (**Figure 12**). This work was published in [30].

C. Assessment of other pathways identified by gene expression profiling of premalignant lesions and invasive squamous lung cancers.

The data from the gene expression profiling of the laser captured airway basal stem cells, premalignant lesions and invasive squamous lung cancer cells suggests that there are multiple metabolic pathways that are important as the basal stem cells undergo hyperplasia and dysplasia and then transfer into invasive cancer cells (**Figure 13**). We are currently assessing these pathways and determining whether they could be successfully targeted in a chemoprevention strategy.

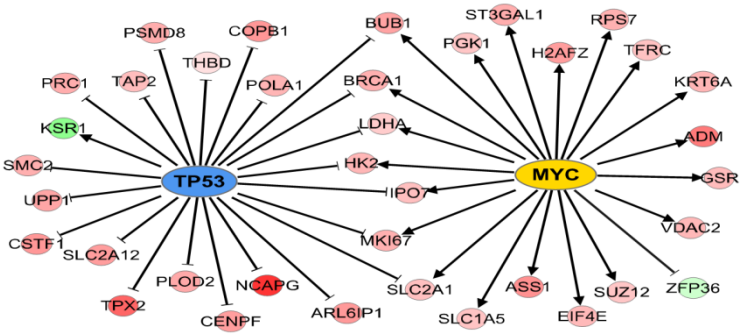


Figure 13. Ingenuity pathway analysis of the P53 and MYC pathway gene expression differences between premalignant lesions and invasive squamous lung cancer. The P53 pathway is predicted to be inhibited, while the MYC pathway is expected to be activated.

D. Analysis of the mutational landscapes of lung premalignancies.

Uncertainty about the clinical behavior of a premalignant lesion can lead to either inappropriate inaction or inappropriate aggressive treatment, either of which can result in harm to the patient. Therefore, in the framework of this DOD funding, we initiated the assessment of the role of somatic mutations that occur within the pulmonary field of cancerization, in early pathogenesis of lung cancer. To this end, we first identified areas of ADC tumor, adenocarcinoma *in situ* (AIS, if present), AAH and histologically normal regions in FFPE tissue blocks from patients who underwent surgical resection of lung adenocarcinoma. We then used LCM to retrieve cells from each of these regions within the same subject (in these preliminary studies, 2-4 areas of each tissue type were isolated and sequenced for each subject). DNA was then extracted from these tissues and the sequencing libraries were constructed and exome enriched. The libraries were sequenced on Illumina HiSeq2000 instrument as 100bp paired-end runs. Most FFPE samples provided high-quality sequencing reads and we obtained >20x — >60x per base mean coverage. Sequencing was repeated for the samples that provided lower than mean 30x coverage.

Next, the sequencing data was analyzed for the presence of somatic mutations. A variant was defined as somatic mutation when $p < 0.1$ for the Fisher test comparing the number of the reference and variant reads in the normal and abnormal (AAH, AIS or ADC) samples. Within each case, position-specific missense and nonsense mutations were compared. Different cases were compared for the mutations at gene-specific level. We define mutations found only in AAH as premalignant, in ADC as malignant and in AAH, AIS and ADC as progression-associated mutations. We identified 367 point mutations affecting amino acid sequences of 362 proteins in more than one abnormal sample. Only two cases out of 9 sequenced in the pilot experiment, were found to bear oncogenic *KRAS* mutations in ADC; in one of these cases the same mutation (at lower read count) was also detected in the AAH lesion, in another case all AAH lesions had the wild type *KRAS*. In other sequenced cases, mutations in *ERBB2* and *FGFR* oncogenes were detected in ADC but not in AAH or AIS samples. The data was compared to the Catalogue of Somatic Mutations in Cancer (COSMIC) database. Among affected proteins, only 3% (i.e. 11 genes) overlapped with the COSMIC cancer census list (**Figure 14**). Interestingly, all of them were identified as common mutations in AAH and at least one ADC sample. Additionally, we compared the recurrently mutations identified in our experiments with The Cancer Genome Atlas (TCGA) database (**Figure 15**). The analysis demonstrated that the majority of the identified mutations were novel, i.e. not included in COSMIC database.

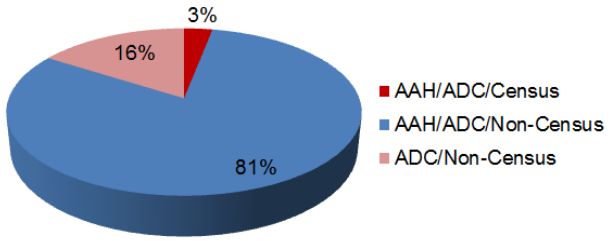


Figure 14. Common point mutations identified among subclones in the same patient. The blue area indicates mutations found in AAH, AIS and ADC within the same patient and not included in COSMIC cancer census list; red area — mutations common in AAH and ADC and included in COSMIC; pink area — mutations unique to ADC and not included in COSMIC.

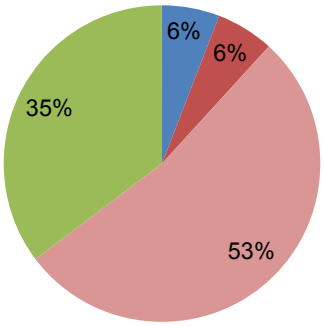


Figure 15. Comparison of the recurrently mutated genes with TCGA. Malignant high: genes mutated in 4 out of 9 patients, low — mutated in 2-3 out of 9 patients.

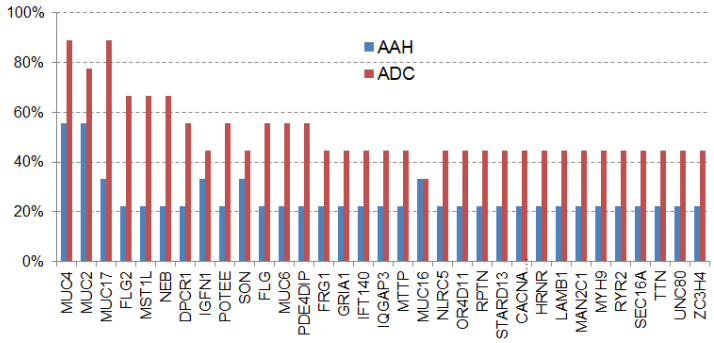


Figure 16. Genes bearing recurrent progression-associated mutations in 4 out of 9 cases.

Further analysis demonstrated that approximately 100 genes bore the progression-associated mutations that were common in at least 2 out of 9 cases sequenced initially. **Figure 16** demonstrates 33 genes that were found to bear progression-associated mutations, common in 4 out of 9 cases (*PDE4DIP* and *MYH9* are included in COSMIC census list). Interestingly, recurring mutations were found in genes involved in calcium signaling and extracellular matrix/receptor interaction (data not shown). These findings demonstrating the role of altered calcium signaling and cell adhesion in early malignant transformation, are novel and have not been reported before. Furthermore, pathways affected by the mutated genes were identified utilizing Gene Ontology and pathways from the KEGG, Biocarta or Reactome databases (**Figure 17**). The observation that few genes mutated in both AAH and ADC are known as key drivers based on either COSMIC or TCGA data, indicates that: a) progression associated mutations might facilitate the malignant transformation by mutated

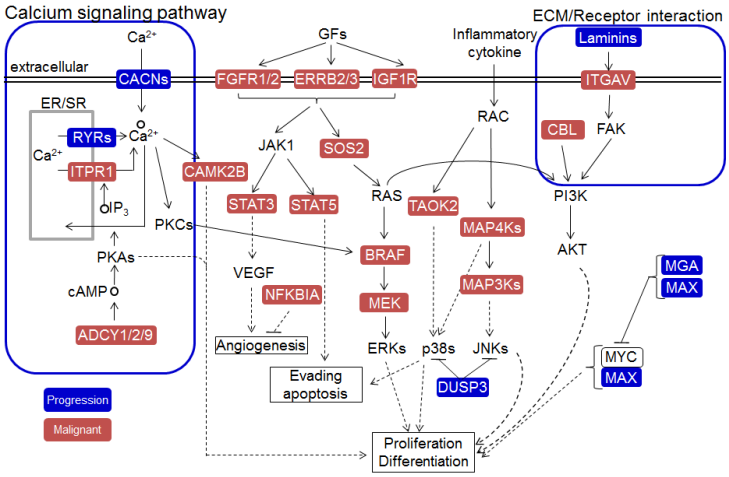


Figure 17. Aberrant pathways enriched by progression-associated and malignant mutants.

key driver(s), or b) a combination of two or more progression associated mutations that are not oncogenic alone, might drive malignant transformation. These hypotheses will be further tested by mapping progression and malignant associated genes into pathway context.

These new data are critically important, given the fact that patients at risk for lung cancer may have subclinical disease for years prior to presentation and diagnosis. Moreover, the proteins encoded by some of mutated genes have recently proven to be tractable targets for new anticancer drug development [31-34]. We anticipate that our results will allow for the development of chemopreventive and/or therapeutic strategies that are effective during the window of opportunity when the disease is still subclinical. Future translation of these results could lead to a paradigm shift in the way in which we address the clinical problem of the individual at high risk for lung cancer.

E. Assessment of the molecular profiles of tumor-initiating stem/progenitor cells from normal airway epithelium, premalignant lesions and cancer.

In collaboration with the Wistuba laboratory, the Massion laboratory has collected bronchial brushings from individuals at risk or with small cell lung cancer. We have collected on 13 individuals with SCLC, isolated total RNA and were able to obtain quality RNA on 11 of them as shown in **Table 4**. The RNA is part of a study to derive a signature for SCLC risk in the field of cancerization and to attempt to identify candidate tumor initiating progenitor stem cells in SCLC.

Table 4. Bronchial brushings from individuals diagnosed with SCLC yet untreated.

project 2967 #	Sample	Barcode	RIN	28s:18s	conc (ng/ul)	yield (ug)	SCLC
95	RNA	522214	7.9	1.3	260	3.9	X
86	RNA	519951	7.6	1.1	28	0.42	X
182	RNA	554960	7.6	1.3	283	4.245	X
177	RNA	515207	7.1	2.1	92	1.38	X
180	RNA	539943	7	1.9	52	0.78	X
68	RNA	524855	6.9	0.9	18	0.27	X
74	RNA	524354	6.7	0.9	247	3.705	X
181	RNA	532271	6.3	2	41	0.615	X
179	RNA	555886	2.4	2.1	68	1.02	X
178	RNA	522995	2.3	1.7	130	1.95	X

F. To increase our understanding of the role of tumor-initiating stem/progenitor cells in the pathogenesis of lung cancer in the "field cancerization" that develops in current and former smokers.

RNA sequencing of laser captured samples to identify genes whose expression is associated with different pathologic stages of lung SCC carcinogenesis (Aim 2B). The first year of this work involved troubleshooting to refine methods for laser capture of normal basal cells (BC), premalignant metaplastic/dysplastic cells, and squamous cell carcinoma (SCC) tumor cells from frozen sections and subsequent RNA isolation. RNA from one subject was hybridized to microarrays, and the expression of control genes behaved as expected, indicating that it was feasible to obtain meaningful results from this material. RNA was then isolated from several more subjects and successfully used to perform high-throughput sequencing (RNA-seq). To identify genes associated with SCC progression, genes with low expression were removed, two different statistical models were used to assign significance to each gene with respect to each pairwise comparison, and the results were intersected to identify genes whose expression changed significantly at an early (normal to premalignant) or late (premalignant to tumor) stage or at both stages (in a stepwise manner). The expression of these genes in all samples from all four patients is illustrated in **Figure 18**.

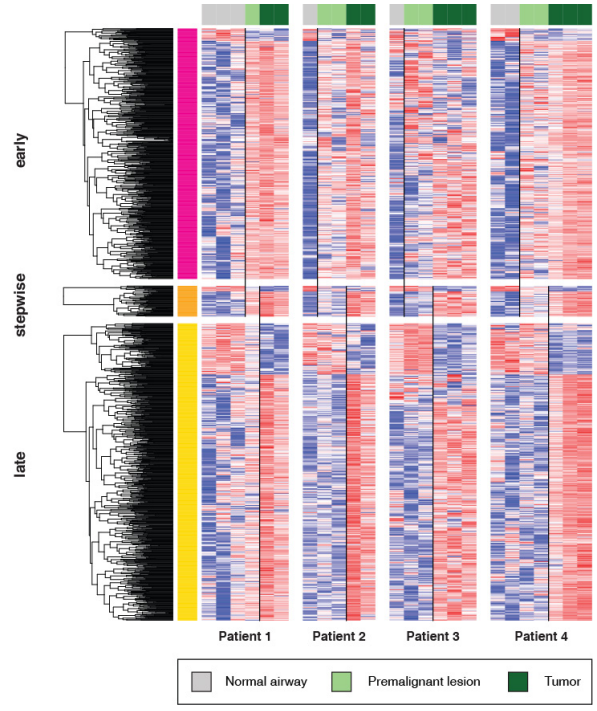


Figure 18. Heatmap of expression of early, late, and stepwise genes. Red and blue indicate expression that is higher or lower than the mean within each patient, respectively. Genes are hierarchically clustered within each group (early, stepwise, late).

Experimental validation of candidate genes with quantitative real-time PCR and immunofluorescence staining (Aim 2B). Three genes (*CEACAM5*, *SLC2A1* and *PTBP3)* with potential roles in lung cancer biology that were up-regulated in premalignant lesions and tumor cells were selected for further validation. qPCR confirmed the upregulation of *CEACAM5* and *SLC2A1* in premalignant cells, and immunofluorescent staining showed that all three genes were undetectable in normal BC but highly expressed in metaplastic and SCC cells. Additionally, the Gene Expression Omnibus (GEO) Profiles tool was used to examine their expression in GEO DataSets relevant to lung SCC carcinogenesis. *SLC2A1* and *PTBP3* were significantly up-regulated in SCC tumors (GDS1312) [35] with respect to paired adjacent normal tissue and in a set of SCC lung tumors relative to lung adenocarcinomas (GDS3627) [36], and *CEACAM5* and *SLC2A1* were significantly up-regulated in bronchoscopic brushings from healthy current smokers relative to healthy never smokers (GDS534) [37].

Computational identification and experimental validation of dysregulated biological mechanisms associated with progression (Aim 2B). Gene Set Enrichment Analysis (GSEA) [38] performed using positionally defined gene sets revealed that the transition from premalignant to tumor is associated with coordinate loss of expression in 3p and gain of expression in 3q26.33-3q29, corresponding to previous reports [39-41]. Ingenuity Pathway Analysis (IPA) was then used to show that the early and late stages were marked by cell cycle progression, and cellular migration and transformation, respectively, and both stages were marked by an increase in cell proliferation and decrease in cell death. IPA was also used to show that genes that are differentially expressed early in carcinogenesis and remain dysregulated in tumor cells are enriched in genes previously reported to be up-regulated by MYC and/or down-regulated by TP53. Immunofluorescent staining was used to confirm that MYC was localized exclusively to the nuclei of premalignant lesions and tumor cells, but not in those of normal BC. More information about this work may be found in a recent publication in Cancer Prevention Research [29].

Specific Aim 3: Test airway-based mRNA and microRNA biomarkers of diagnosing lung cancer in current and former smokers at high risk for lung cancer in minimally invasive sites.

Mapping the airway-wide molecular field of injury in smokers with lung cancer.
In Years 03 and 04, we studied the molecular spatial map of field effects that transverse the normal-appearing bronchus adjacent to tumors up to the relatively distant nasal epithelium. We surmised that this analysis would aid in identification of shared genomic changes between the field and lung cancer and that extend to compartments (e.g. nasal) in the field cancerization that can be readily accessible for biomarker analysis in screening and clinical settings. Samples (n=254) from patients with (n=28) and without (n=9) lung cancer that were collected from all partnering institutions (14 cancer cases from MD Anderson) were processed for global expression profiling at MD Anderson Cancer center using the Human Gene 2.0 ST platform (Affymetrix) and data analysis was performed in collaboration with BU (Partnering PI, Dr. Avrum Spira) (please see report by Dr. Spira). We identified profiles differentially expressed between NSCLCs and benign nodules. 47 genes were identified to be differentially expressed in a gradient like manner in the field of injury, i.e. genes that increase or decrease in expression with increasing distance from the primary lung tumor (**Figure 19**, left heat map). Gene expression analysis also identified 106 genes to be significantly differentially expressed in the airways of NSCLC patients compared to

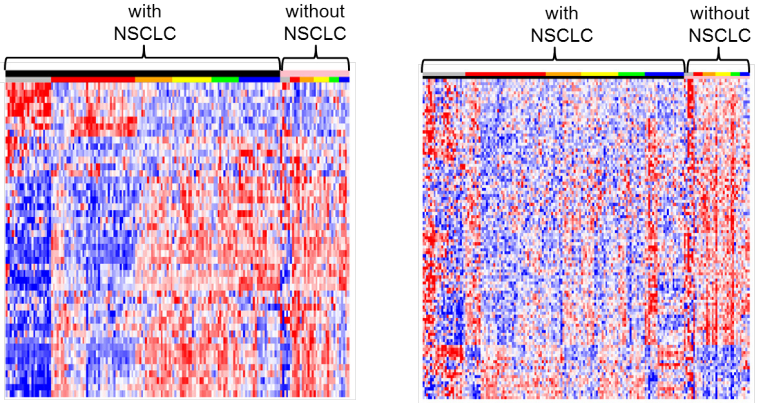


Figure 19. Airway-wide gene expression patterns in the field of injury. Gene expression profiling identified two distinct patterns of gene expression in the airway field of injury: 47 genes identified to be differentially expressed in a gradient manner in the field of injury of NSCLC patients and not in patients with benign disease (left heatmap) and 106 genes that were found to be differentially expressed between NSCLC patients and smokers with benign disease and that were persistent throughout the airway field of injury.

smokers with benign disease and that were persistently expressed (non-gradient) throughout the airway field of injury (**Figure 19**, right heatmap). Importantly, these gradient and persistent genes were validated in independent samples and cohorts (please see report by Dr. Avrum Spira). These data point to field of injury profiles that embody the nearby lung tumor and inform of lung cancer pathogenesis (gradient field of injury) and field markers that are more likely to be diagnostic of lung cancer in minimally invasive sites in the clinical setting (persistent genes). These findings have been presented by Kusko et al. in the past 2014 American Association for Cancer Research annual meeting [42] and are currently being prepared as a manuscript for publication.

4. KEY RESEARCH ACCOMPLISHMENTS

Year 01

- Identified that gene expression is modulated in a site- and a time-dependent manner in the bronchial epithelium of early stage lung cancer patients.
- Identified several pathways preferentially activated in the airway adjacent to tumors in patients with lung cancer, including those mediated by PI3K, NF-kB and ERK1/2.
- Completed the collection and field cancerization gene expression analysis of 23 patients (n=226 samples) with lung tumors using samples obtained from lobectomy specimens.

Year 02

- Identified aberrant activation of canonical oncogenes in the molecular field of injury of early stage NSCLC patients including phosphorylated AKT and ERK kinases.
- Completed the analysis and characterization of the molecular localized field cancerization using 194 samples comprised of tumor, normal lung and airway samples from 20 NSCLC cases.
- Derived field cancerization expression signatures, comprised of genes concordantly and significantly differentially expressed between tumors and airways compared to matched normal lung tissue, and pertinent to NSCLC, lung adenocarcinomas and SCCs.
- In collaboration with BU (Partnering institution), demonstrated that molecular profiles in the localized field cancerization are, at least in part, relevant to the molecular field of injury and contain markers significantly and concordantly different between airways of patients with and without lung cancer
- Identified profiles that were significantly progressively and differentially expressed by distance from corresponding lung tumors and concordantly modulated between tumors and paired uninvolved normal lung tissues, pinpointing to their probable roles in pathogenesis.
- Confirmed the differential expression of several site-dependent and –independent markers identified by microarray analysis of the field cancerization at the RNA and protein level by QRTPCR and immunohistochemical analysis, respectively.
- In collaboration with the Partnering PIs and Initiating PI of this award, began to analyze by RNA-Seq at BU (20 cases, 156 samples) and comprehensive microarray profiling at MD Anderson (28 cases, 183 samples) the molecular field of injury in NSCLC patients and cancer-fee individuals.

Year 03

- Characterized the transcriptomic architecture of the adjacent airway field cancerization in early-stage non-small cell lung cancer. These analyses demonstrated that the adjacent airway field of cancerization is comprised of markers that can identify lung cancer among smokers as well as gradient and localized site-dependent expression patterns that recapitulate NSCLC profiles. These findings have been submitted recently for publication and are under revision.
- Demonstrated for the first time that the field cancerization putative oncogene, *LAPTM4B*, is a positive mediator of the lung cancer cell malignant phenotype evidenced by its promotion of anchorage-independent colony formation in soft agar.
- Studied the mRNA expression of *LAPTM4B* in a large series of NSCLC histological tissue specimens for the first time by *in situ* hybridization. This analysis revealed that *LAPTM4B* expression is significantly positively associated with smoking and worse overall survival.
- Demonstrated that the field cancerization marker *LAPTM4B* protects lung cancer cells from serum deprivation-induced growth inhibition and promotes the autophagy response following serum deprivation.
- Revealed that *LAPTM4B* is a novel positive regulator of NRF2 transcription factor in lung cancer cells.

- In collaboration with the Partnering PIs and Initiating PI of this grant, performed microarray profiling at MD Anderson of 254 field cancerization samples from 28 cases with lung cancer and 9 cases with benign disease to begin to characterize the molecular spatial map of field effects that transverse the bronchus adjacent to tumors up to the nasal epithelium. This novel analysis demonstrated that the molecular map of the field of injury/cancerization, in patients with lung cancer, is comprised of pathways and gene sets, whose enrichment in the field decreases with larger distance from the tumor as well as those that persist up to the nasal epithelium.
- Identified specific pathways that are involved in stepwise progression to SCC.
- Found that metabolic changes in the airway basal stem cells are critical for stepwise progression

Year 04

- Began to understand genomic aberrations such as point mutations (i.e. single nucleotide polymorphisms, small insertions and deletions) and copy number alterations (i.e. amplification, deletion, loss-of-heterozygosity, allelic imbalance) in the airway field cancerization using deep sequencing and SNP array analyses. These findings in the last year of the DoD grant is the foundation for current and future studies to investigate the genomic architecture of the airway field of cancerization in lung cancer.
- Completed series of *in vitro* experiments demonstrating the interplay between LAPTM4B and NRF2 in mediating stress responses that are vital for the survival of lung cancer cells.
- In collaboration with the Partnering PIs and Initiating PI of this grant, completed microarray analysis of 254 field cancerization samples from 28 cases with lung cancer and 9 cases with benign disease to characterize in depth the molecular field of injury and identify field of injury markers that inform of lung cancer pathogenesis and that are diagnostic of lung cancer in minimally invasive sites.
- Compared the mutational landscapes of the primary lung tumors, premalignant lesions and adjacent normal tissues, and identified novel genes associated with malignant transformation.
- Found that reactive oxygen species metabolism plays a vital role in homeostasis of the airway basal stem cells and that perturbing this homeostatic mechanism results in premalignant lesions.
- Found that Nrf2 and Notch1 act downstream of ROS in this homeostatic regulation that prevents excessive self-renewal and promotes mucociliary differentiation.

5. CONCLUSION

During this DOD program, we 1) understood how the molecular field of injury evolves spatiotemporally following curative surgery in smoker early-stage NSCLC patients; 2) characterized the transcriptomic architecture of the adjacent airway field cancerization in early-stage NSCLC; 3) identified gradient profiles in the localized field cancerization that highly embody the nearby lung tumors; 4) demonstrated that the molecular adjacent field cancerization extends to relatively less invasive large airways and harbors markers that can detect lung cancer in smokers that are suspect of the malignancy; 5) validated the expression of a novel field cancerization marker, *LAPTM4B*, in airways and tumors in NSCLC patients; 6) revealed a novel interplay between *LAPTM4B* and the *NRF2* transcription in mediating the NRF2-stress response and pathway for ensuing lung cancer cell growth and survival; 7) began to understand genomic aberrations (mutations and copy number alterations) in the airway field of cancerization and 8) identified field of injury/cancerization markers that are specific to NSCLC vs benign disease, inform of NSCLC pathogenesis and can detect lung cancer when assessed in minimally invasive sites in the lung.

Also, our work has provided most comprehensive understanding to date of progression of airway basal stem cells to premalignant lesions and squamous lung carcinogenesis. The pathways identified, such as the ROS pathway, provide valuable insight into the biology of stepwise progression to SCC in current and former smokers. In the future we plan to identify the drivers that are critical in taking premalignant lesions to the point of invasive SCC. This will hopefully lead to the development of novel targeted therapies to prevent and treat SCC.

We determined that bronchoscopy specimens, such as brushings, can be used successfully for shotgun proteomic profiling. Thus, utilizing bronchial brushings, we successfully demonstrated by shotgun proteomics that subjects at increased risk for lung cancer have a characteristic differential protein expression signature. Based on these findings, we identified a signature of lung cancer risk assessment biomarkers that may provide the basis of patient selection for surveillance programs and chemoprevention. Pathway enrichment analysis as well as *in vitro* data, suggests possible metabolic reprogramming in the airway epithelium of high risk individuals.

6. PUBLICATIONS, ABSTRACTS, AND PRESENTATIONS

Abstracts:

- Kadara H, Saintigny P, Fan Y, Chow CW, Chu ZM, Lang W, Behrens C, Gold K, Liu D, Lee JJ, Mao L, Kim ES, Hong WK, Wistuba II. Gene expression analysis of field of cancerization in early stage NSCLC patients towards development of biomarkers for personalized prevention. Proceedings of the 102nd Annual Meeting of the American Association for Cancer Research; 2011 Apr 2-6; Orlando, Florida. Philadelphia (PA): AACR; 2011. Abstract #3674.
- Kadara H, Fujimoto J, Yoo SY, Garcia M, Kabbout M, Basey A, Wang J, Coombes KR, Kim ES, Hong WK, Kalhor N, Moran C, Wistuba II. Gene expression profiling of lung tumors and matched normal airways reveals common and disparate aberrant pathways in squamous cell carcinoma and adenocarcinoma development and potential targets for chemoprevention in early stage lung cancer patients. Proceedings of the 103rd Annual Meeting of the American Association for Cancer Research; 2012 Mar 31- Apr 4; Chicago, Illinois. Philadelphia (PA): AACR; 2012. Abstract #1721.
- Maki Y, Fujimoto J, Yoo SY, Gower A, Shen L, Garcia MM, Kabbout M, Chow CW, Hong WK, Kalhor N, Wang J, Moran C, Spira A, Coombes KR, Wistuba II, Kadara H. Transcriptomic architecture of the airway field cancerization in early-stage non-small cell lung cancer. 104th Annual American Association for Cancer Research (AACR) meeting, April 6 - April 10 2013, Washington, D.C. Abstract # 2367.
- Maki Y, Fujimoto J, Yoo SY, Garcia M, Gower A, Shen Li, chow CW, Behrens C, Kalhor N, Moran C, Wang J, Spira A, Coombes KR, Wistuba II, Kadara H. Transcriptomic architecture of the field of cancerization in the adjacent normal-appearing airway: Early mechanisms in lung carcinogenesis. Proceedings of the 105th Annual Meeting of the American Association for Cancer research; April 5-9 2014; San Diego, CA. Abstract #1597.
- Kusko R, Anderlind C, Wang G, Zhang S, Wallace WD, Wasler T, Ebright M, Garcia MM, Eisenberg R, Lee G, Liu G, Elashoff D, Kalhor N, Moran C, Mehran R, Fujimoto F, Massion PP, Dubinett S, Wistuba II, Lenburg M, Kadara H*, Spira A*. Mapping the airway-wide molecular field of injury in smokers with lung cancer. Proceedings of the 105th Annual Meeting of the American Association for Cancer research; April 5-9 2014; San Diego, CA. Abstract #2352.
- Ooi AT, Gower AC, Zhang K, Vick J, Hong LS, Fishbein M, Nagao B, Wallace WD, Elashoff DA, Dubinett S, Lenburg M, Spira A, Gomperts BN. Gene expression alterations in premalignant lesions from the airways of patients with lung squamous cell carcinomas. Platform presentation and travel award. 104th Annual American Association for Cancer Research (AACR) meeting, April 6 - April 10 2013, Washington, D.C.
- Bisht B, Paul M, Darmawan D, Chon A, Chiu R, Alva JM and Gomperts BN. Redox regulation of airway epithelial stem cell maintenance and repair. ISSCR poster presentation. Boston June 2013.
- Krysan K, Walser T, Kusko R, Gardner BK, Spira AP, Lenburg ME, Brothers JF 2nd2, O'Hearn J, Reckamp KL, Mao JT, Purcell C, Lee JM, Garon EB, Dubinett SM. PGE2-dependent upregulation of the miR-19-72 cluster contributes to apoptosis resistance in NSCLC. Lung Cancer SPORE Workshop, July 16-18, 2012, Pittsburgh, PA.
- Park SJ, Walser TC, Tran LM, Perdomo C, Wang T, Hong LS, Pagano PC, Liclican EL, Krysan K, Larsen JE, Fishbein MC, Minna JD, Lenburg ME, Spira A, Dubinett SM. The impact of e-cigarettes exposure on pulmonary epithelium gene expression and transformation. Lung SPORE Workshop, July 25-26 2014, Pittsburgh, PA.
- Pavel A, Campbell J, Liu G, Zhang S, Liu H, Dubinett S, Elashoff D, Porta K, Whitney D, Lenburg M, Spira A. Biomarker development for lung cancer diagnosis using integrative microRNA and gene expression networks. Proceedings of the American Association for Cancer Research April; 5-9 2014; San Diego, CA. Abstract #1485.
- Kusko R, Anderlind C, Wang G, Zhang S, Wallace WD, Wasler T, Ebright M, Garcia MM, Eisenberg R, Lee G, Liu G, Elashoff D, Kalhor N, Moran C, Mehran R, Fujimoto J, Massion PP, Dubinett S, Wistuba I, Lenburg M, Kadara H, Spira A. Mapping the airway-wide molecular field of injury in smokers with lung cancer. Proceedings of the American Association for Cancer Research; 5-9 2014; San Diego, CA. Abstract #2352.
- Rahman SMJ, Li M, Zou Y, Zimmerman ML, Lu P, Slebos R, Zhang B, Shyr Y, Liebler D and Massion PP. Proteomic analysis of the bronchial epithelium to develop a molecular signature of risk for lung cancer. IASLC 15th World Conference on Lung Cancer, Sydney, Australia, October, 2013. (ID 2008)

- Jun Qian, Xiangming Ji, Yong Zou, Megan D. Hoeksema, Heidi Chen and Pierre P. Massion. Overexpression of miRNA-1224-5p, 1248 and 944 on chromosome 3q26-29 in squamous cell carcinoma of the lung. 1. 2014 AACR meeting abstract (#4989).
- Krysan K, Walser T, Seki A, Tran L, Spira A, Fishbein M, Dubinett S. The mutational landscape of pulmonary premalignancy in the context of lung adenocarcinoma. Lung Cancer SPORE Workshop, July 25-26 2014, Pittsburgh, PA.

Manuscripts:

- Kadara H, Kabbout M, Wistuba II. Pulmonary adenocarcinoma: a renewed entity in 2011. *Respirology*. 7(5):833-40 2012.
- Kadara H, Wistuba II. Field cancerization in NSCLC: implications in disease pathogenesis. *Proceedings of the American Thoracic Society*. 9(2):38-42 2012.
- Fujimoto J*, Kadara H* (* equal contributing first authors), Garcia MM, Kabbout M, Behrens C, Liu DD, Lee JJ, Solis LM, Kim ES, Kalhor N, Moran C, Shalafkhaneh A, Lotan R, Wistuba II. G-protein coupled receptor family C group 5 member A (GPRC5A) expression is decreased in the adjacent field and normal bronchial epithelia of patients with chronic obstructive pulmonary disease and non-small cell lung cancer. *Journal of Thoracic Oncology*. 7(12):1747-54 (2012).
- Kadara H, Shen L, Fujimoto J, Saintigny P, Chow CW, Lang W, Chu Z, Garcia M, Kabbout M, Fan YH, Behrens C, Liu D, Mao L, Lee JJ, Gold KA, Wang J, Coombes K, Kim ES, Hong WK, Wistuba II. Characterizing the molecular spatial and temporal field of injury in early stage smoker non-small cell lung cancer patients after definitive surgery by expression profiling. *Cancer Prevention Research*. 6(1):8-17 2013.
- Kadara H, Fujimoto J, Yoo SY, Maki Y, Gower AC, Kabbout M, Garcia MM, Chow CW, Chu Z, Mendoza G, Shen L, Kalhor N, Hong WK, Moran C, Wang J, Spira A, Coombes KR, Wistuba II. Transcriptomic architecture of the adjacent airway field cancerization in non-small cell lung cancer. *Journal of the National Cancer Institute*. 2014 106(3), in press.
- Gomperts BN, Spira A, Massion PP, Walser TC, Wistuba, II, Minna JD, Dubinett SM. Evolving concepts in lung carcinogenesis. *Seminars in Respiratory and Critical Care Medicine* 2011 32(1):32-43. PMCID: 3423907.
- Ooi AT, Gower AC, Zhang KX, Vick JL, Hong L, Nagao B, et al. Molecular profiling of premalignant lesions in lung squamous cell carcinomas identifies mechanisms involved in stepwise carcinogenesis. *Cancer Prevention Research*. 2014;7(5):487-95. PMCID: 4059064.
- Paul MK, Bisht B, Darmawan DO, Chiou R, Ha VL, Wallace WD, et al. Dynamic changes in intracellular ROS levels regulate airway basal stem cell homeostasis through Nrf2-dependent Notch signaling. Cell Stem Cell. 2014;15(2):199-214. PMCID: 4127166.
- Gomperts BN, Walser TC, Spira A, Dubinett SM. Enriching the molecular definition of the airway "field of cancerization:" establishing new paradigms for the patient at risk for lung cancer. *Cancer Prevention Research (Phila)*. 2013 6(1):4-7.
- Dubinett SM, Spira A. he challenge and opportunity of targeted lung cancer chemoprevention. *Journal of Clinical Oncology* 2013 31(33):4169-71.
- Harris FT, Rahman SMJ, Hassanein M, Qian J, Hoeksema M, Chen H, Eisenberg R, Chaurand P, Caprioli RM, Shiota M and Massion PP. Acyl-Coenzyme A Binding Protein Regulates Beta Oxidation Required for Growth and Survival of Non-Small Cell Lung Cancer. *Can Prev Res*. 7:748-57, 2014.
- Grant JL, Fishbein MC, Hong LS, Krysan K, Minna JD, Shay JW, Walser TC, Dubinett SM. (2014) A novel molecular pathway for Snail-dependent, SPARC-mediated invasion in non-small cell lung cancer pathogenesis. *Cancer Prev Res.*, 7, 150-160. PMID: 24253315.
- Krysan K, Cui X, Gardner BK, Reckamp KL, Wang X, Hong L, Walser TC, Rodriguez NL, Pagano PC, Garon EB, Brothers JF II, Elashoff D, Lee JM, Spira AE, Sharma S, Fishbein MC, Dubinett SM. Elevated neutrophil gelatinase-associated lipocalin contributes to erlotinib resistance in non-small cell lung cancer. *American Journal of Translational Research* 2013 5(5):481-96. PMCID: PMC3745436.
- Krysan K, Kusko R, Grogan T, O'Hearn J, Reckamp K, Walser TC, Garon EB, Lenburg ME, Sharma S, Spira AE, Elashoff D, Dubinett SM. (2014) PGE2-dependent upregulation of the miR-19-72 cluster contributes to apoptosis resistance in NSCLC. *Mol Cancer Res.*, 12, 765-774. PMID: 24469837.

7. INVENTIONS, PATENTS AND LICENSES

Nothing to report.

8. REPORTABLE OUTCOMES

- Pease see the list of abstracts and publications above for the complete list of reportable outcomes.
- List of genes differentially expressed during stepwise progression to SCC deposited in GEO.
- Developed a protocol for massively parallel sequencing of RNA isolated from premalignant lesions from FFPE archived tissues by laser capture microdissection.
- Developed a protocol for whole exome DNA sequencing of cells of interest in lung cancer continuum, laser capture microdissected from FFPE archived tissues.
- Developed a protein signature that is able to differentiate subjects at risk for lung cancer and lung cancer patients.

9. OTHER ACHIEVEMENTS

Aik Ooi received employment at Fluidigm company based on his research experience on this project.

10. REFERENCES

1. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F*, et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nat Med 2007;13:361-6.
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin 2013;63:11-30.
3. Services USDoHaH. The Health Consequences of Smoking. A Report of the U.S. Surgeon General. 2004.
4. National Lung Screening Trial Research T, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD*, et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365:395-409.
5. Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR*, et al.* Results of the two incidence screenings in the National Lung Screening Trial. N Engl J Med 2013;369:920-31.
6. Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X*, et al.* Effects of cigarette smoke on the human airway epithelial cell transcriptome. Proc Natl Acad Sci U S A 2004;101:10143-8.
7. Sridhar S, Schembri F, Zeskind J, Shah V, Gustafson AM, Steiling K*, et al.* Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. BMC Genomics 2008;9:259.
8. Beane J, Sebastiani P, Whitfield TH, Steiling K, Dumas YM, Lenburg ME*, et al.* A prediction model for lung cancer diagnosis that integrates genomic and clinical features. Cancer Prev Res (Phila Pa) 2008;1:56-64.
9. Bhutani M, Pathak AK, Fan YH, Liu DD, Lee JJ, Tang H*, et al.* Oral epithelium as a surrogate tissue for assessing smoking-induced molecular alterations in the lungs. Cancer Prev Res (Phila Pa) 2008;1:39-44.
10. Kadara H, Kabbout M, Wistuba, II. Pulmonary adenocarcinoma: a renewed entity in 2011. Respirology 2012;17:50-65.
11. Kadara H, Wistuba, II. Field cancerization in non-small cell lung cancer: implications in disease pathogenesis. Proc Am Thorac Soc 2012;9:38-42.
12. Kadara H, Shen L, Fujimoto J, Saintigny P, Chow CW, Lang W*, et al.* Characterizing the molecular spatial and temporal field of injury in early-stage smoker non-small cell lung cancer patients after definitive surgery by expression profiling. Cancer Prev Res (Phila) 2013;6:8-17.
13. Gomperts BN, Walser TC, Spira A, Dubinett SM. Enriching the molecular definition of the airway "field of cancerization:" establishing new paradigms for the patient at risk for lung cancer. Cancer Prev Res (Phila) 2013;6:4-7.
14. Kadara H, Fujimoto J, Yoo SY, Maki Y, Gower AC, Kabbout M*, et al.* Transcriptomic architecture of the adjacent airway field cancerization in non-small cell lung cancer. J Natl Cancer Inst 2014;106:dju004.
15. Chen Y, Deng J, Fujimoto J, Kadara H, Men T, Lotan D*, et al.* Gprc5a deletion enhances the transformed phenotype in normal and malignant lung epithelial cells by eliciting persistent Stat3 signaling induced by autocrine leukemia inhibitory factor. Cancer Res 2010;70:8917-26.
16. Deng J, Fujimoto J, Ye XF, Men TY, Van Pelt CS, Chen YL*, et al.* Knockout of the tumor suppressor gene Gprc5a in mice leads to NF-kappaB activation in airway epithelium and promotes lung inflammation and tumorigenesis. Cancer Prev Res (Phila) 2010;3:424-37.

17. Fujimoto J, Kadara H, Men T, van Pelt C, Lotan D, Lotan R. Comparative functional genomics analysis of NNK tobacco-carcinogen induced lung adenocarcinoma development in Gprc5a-knockout mice. PLoS One 2010;5:e11847.
18. Kadara H, Fujimoto J, Men T, Ye X, Lotan D, Lee JS*, et al.* A Gprc5a tumor suppressor loss of expression signature is conserved, prevalent, and associated with survival in human lung adenocarcinomas. Neoplasia 2010;12:499-505.
19. Tao Q, Fujimoto J, Men T, Ye X, Deng J, Lacroix L*, et al.* Identification of the retinoic acid-inducible Gprc5a as a new lung tumor suppressor gene. J Natl Cancer Inst 2007;99:1668-82.
20. Fujimoto J, Kadara H, Garcia MM, Kabbout M, Behrens C, Liu DD*, et al.* G-protein coupled receptor family C, group 5, member A (GPRC5A) expression is decreased in the adjacent field and normal bronchial epithelia of patients with chronic obstructive pulmonary disease and non-small-cell lung cancer. J Thorac Oncol 2012;7:1747-54.
21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A*, et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.
22. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C*, et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213-9.
23. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 2006;78:629-44.
24. Scheet P, Stephens M. Linkage disequilibrium-based quality control for large-scale genetic studies. PLoS Genet 2008;4:e1000147.
25. Vattathil S, Scheet P. Haplotype-based profiling of subtle allelic imbalance with SNP arrays. Genome Res 2013;23:152-8.
26. Li Y, Zhang Q, Tian R, Wang Q, Zhao JJ, Iglehart JD*, et al.* Lysosomal transmembrane protein LAPTM4B promotes autophagy and tolerance to metabolic stress in cancer cells. Cancer Res 2011;71:7481-9.
27. Li Y, Zou L, Li Q, Haibe-Kains B, Tian R, Desmedt C*, et al.* Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. Nat Med 2010;16:214-8.
28. Maki Y, Fujimoto J, Yoo S-Y, Garcia M, Gower A, Shen L*, et al.* Abstract 1597: Transcriptomic architecture of the field of cancerization in the adjacent normal-appearing airway: Early mechanisms in lung carcinogenesis. Cancer Res 2014;74:1597.
29. Ooi AT, Gower AC, Zhang KX, Vick JL, Hong L, Nagao B*, et al.* Molecular profiling of premalignant lesions in lung squamous cell carcinomas identifies mechanisms involved in stepwise carcinogenesis. Cancer Prev Res (Phila) 2014;7:487-95.
30. Paul MK, Bisht B, Darmawan DO, Chiou R, Ha VL, Wallace WD*, et al.* Dynamic changes in intracellular ROS levels regulate airway basal stem cell homeostasis through Nrf2-dependent Notch signaling. Cell Stem Cell 2014;15:199-214.
31. Bronte G, Rizzo S, La Paglia L, Adamo V, Siragusa S, Ficorella C*, et al.* Driver mutations and differential sensitivity to targeted therapies: a new approach to the treatment of lung adenocarcinoma. Cancer Treat Rev 2010;36 Suppl 3:S21-9.
32. Pao W, Kris MG, Iafrate AJ, Ladanyi M, Janne PA, Wistuba, II*, et al.* Integration of molecular profiling into the lung cancer clinic. Clin Cancer Res 2009;15:5317-22.
33. The Cancer Genome Atlas Research N, Genome sequencing centres: Broad I, Hammerman PS, Lawrence MS, Voet D, Jing R*, et al.* Comprehensive genomic characterization of squamous cell lung cancers. Nature 2012.
34. Oxnard GR, Binder A, Janne PA. New targetable oncogenes in non-small-cell lung cancer. J Clin Oncol 2013;31:1097-104.
35. Wachi S, Yoneda K, Wu R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. Bioinformatics 2005;21:4205-8.
36. Kuner R, Muley T, Meister M, Ruschhaupt M, Buness A, Xu EC*, et al.* Global gene expression analysis reveals specific patterns of cell junctions in non-small cell lung cancer subtypes. Lung Cancer 2009;63:32-8.
37. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. Genome Biol 2007;8:R201.
38. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.

39. Partridge M, Kiguwa S, Langdon JD. Frequent deletion of chromosome 3p in oral squamous cell carcinoma. Eur J Cancer B Oral Oncol 1994;30B:248-51.
40. Brunelli M, Bria E, Nottegar A, Cingarlini S, Simionato F, Calio A*, et al.* True 3q chromosomal amplification in squamous cell lung carcinoma by FISH and aCGH molecular analysis: impact on targeted drugs. PLoS One 2012;7:e49689.
41. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG*, et al.* SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet 2009;41:1238-42.
42. Kusko R, Anderlind C, Wang G, Zhang S, Wallace WD, Wasler T*, et al.* Abstract 2352: Mapping the airway-wide molecular field of injury in smokers with lung cancer. Cancer Res 2014;74:2352.

11. **APPENDICES**

Cancer Research
cancerres.aacrjournals.org

doi: 10.1158/1538-7445.AM2011-3674
Cancer Res April 15, 2011 71; 3674

Abstract 3674: Gene expression analysis of field of cancerization in early stage NSCLC patients towards development of biomarkers for personalized prevention

Humam Kadara[1], Pierre Saintigny[1], Youhong Fan[1], Chi-Wan Chow[1], ZuoMing Chu[1], Wenhua Lang[1], Carmen Behrens[1], Kathryn Gold[1], Diane Liu[1], J. Jack Lee[1], Li Mao[2], Edward S. Kim[1], Waun K. Hong[1], and Ignacio I. Wistuba[1]

- Author Affiliations

[1]UT MD Anderson Cancer Center, Houston, TX
[2]University of Maryland, Baltimore, MD.

Proceedings: AACR 102nd Annual Meeting 2011– Apr 2-6, 2011; Orlando, FL

Abstract

Background: The identification of early stage non-small cell lung cancer (ES NSCLC) patients (pts) at higher risk for recurrence or second primary tumor (SPT) development is vital to personalizing prevention and therapy. We sought to decipher spatial and temporal patterns of gene expression in the airway field of ever-smoker ES NSCLC pts to better understand lung cancer pathogenesis and predict recurrence or SPT development.

Methods: Pts on the prospective Vanguard study had definitively treated ES (I/II) NSCLC, were current/former smokers, and had bronchoscopies with brushings obtained from the main carina (MC) at baseline, 12, and 24 months following resective surgery and from different anatomical regions at baseline. Expression profiling is ongoing for all eligible pts (41 pts, 326 samples). To query temporal and spatial airway expression profiles, two sets of six pts were selected based on complete processed time point and baseline airway site (3 different sites per pt) arrays (Affymetrix Human Gene 1.0 ST), respectively. Temporally and spatially differentially expressed genes were independently identified based on a $p<0.01$ of a univariate t-test with estimation of the false discovery rate (FDR), studied by hierarchical clustering and principal component analysis (PCA), and functionally analyzed using network analysis.

Results: 871 gene features were differentially expressed among MCs of six NSCLC pts at baseline, 12 and 24 months and were shown to separately group the MCs as evident in both cluster and PC analyses. Moreover, pathways analysis of the temporally modulated genes showed that a gene-network mediated by extracellular regulated kinase (ERK1/2) was most significantly elevated ($p<0.001$) in function between MCs at 24 months versus baseline. 763 and 931 gene features were differentially expressed between MCs and adjacent-to-resected tumors (ADJ) airways and between MC, ADJ and non-adjacent (distant-to-resected tumor) (NON-ADJ) airways, respectively. Moreover, pathways analysis of the spatially modulated genes revealed that gene-networks mediated by nuclear factor-κB (NF-κB) and ERK1/2-mediated were most significantly elevated ($p<0.001$) in function in ADJ airway samples versus MCs. Furthermore, PCA revealed that while ADJ airway samples grouped separately and closely together, one MC and 3 NON-ADJ airway samples resided closely with ADJ samples, which were then found to originate from 3 pts with evidence of recurrence, SPT or suspicion of recurrence.

Conclusions: Our findings highlight expression signatures and pathways (ERK1/2 and NF-κB) in a "cancerization field" that may drive lung cancer pathogenesis and be associated with recurrence or SPT development in ES NSCLC pts and thus useful for derivation of biomarkers to guide personalized prevention strategies. Supported by DoD grants W81XWH-04-1-0142 and W81XWH-10-1-1007.

Citation Format: {Authors}. {Abstract title} [abstract]. In: Proceedings of the 102nd Annual Meeting of the American Association for Cancer Research; 2011 Apr 2-6; Orlando, FL. Philadelphia (PA): AACR; Cancer Res 2011;71(8 Suppl):Abstract nr 3674. doi:10.1158/1538-7445.AM2011-3674

Cancer Research
cancerres.aacrjournals.org

doi: 10.1158/1538-7445.AM2012-1721
Cancer Res April 15, 2012 72; 1721

Abstract 1721: Gene expression profiling of lung tumors and matched normal airways reveals common and disparate aberrant pathways in squamous cell carcinoma and adenocarcinoma development and potential markers for detection and targets for chemoprevention in early stage lung cancer

Humam Kadara[1], **Junya Fujimoto**[1], **Suk Young Yoo**[1], **Melinda M. Garcia**[1], **Mohamed Kabbout**[1], **Annette L. Basey**[1], **Jing Wang**[1], **Kevin R. Coombes**[1], **Edward S. Kim**[1], **Waun Ki Hong**[1], **Neda Kalhor**[1], **Cesar Moran**[1], and **Ignacio I. Wistuba**[1]

Author Affiliations

[1]UT M.D. Anderson Cancer Center, Houston, TX

Proceedings: AACR 103rd Annual Meeting 2012-- Mar 31-Apr 4, 2012; Chicago, IL

Abstract

Detection and chemoprevention of non-small cell lung cancer (NSCLC) have been unsuccessful in part due to our limited knowledge of the pathogenesis of the disease. It has been suggested that normal airway epithelia share molecular abnormalities with tumors and may serve as progenitors for lung malignancies. We sought to analyze molecular profiles of lung adenocarcinomas and squamous cell carcinomas (SCCs), two major subtypes of NSCLC, and matched normal airways to elucidate aberrant expression patterns in early phases of lung tumorigenesis. All specimens were obtained from primary early stage NSCLC consented patients who did not receive neoadjuvant therapy (n=20). We profiled RNA isolated from tumors and normal lung as well as from brushings of multiple matched airways that were histologically confirmed to lack neoplastic or preneoplastic cells (n=194). Expression signatures signifying genes significantly and concurrently differentially expressed between both tumors and airways compared to normal lung tissue (tumor-airway-normal/TAN signatures) were then derived independently for SCC and smoker adenocarcinoma cases and were comprised of 1,803 and 1,938 genes, respectively. The TAN signatures effectively clustered tumor and airways from normal lung samples ($p<0.001$). Further analysis showed that a subset of the genes separated SCC- and adenocarcinoma-adjacent airways. Moreover, pathways and gene-network analysis using Ingenuity pathways software highlighted similarities and differences in pathway modulation between airway epithelial fields of SCCs and adenocarcinomas. Embryonic stem cell and eicosanoid signaling pathways were most significantly modulated among those common to both TAN signatures ($p<0.001$). Retinoic acid receptor and stem cell signaling pathways mediated by NANOG and lineage oncogene SOX2 were most significantly modulated in the SCC TAN signature, whereas NF-kB and PTEN signaling pathways were most prevalent in the adenocarcinoma TAN airway signature (all $p<0.001$). Gene networks mediated by lineage oncogene NKX2-1/TITF-1 down-regulation and TP63 up-regulation, and networks mediated by increased expression of the MET and ERBB2 oncogenes were predominantly functionally modulated in the SCC and adenocarcinoma TAN signatures, respectively. Quantitative PCR analysis confirmed up-regulation of MET in adenocarcinomas and normal airways compared to normal lung. These findings highlight expression patterns and pathways that are deregulated differentially in the pathogenesis of lung adenocarcinomas and SCCs and thus offer detection markers as well as therapeutic targets to guide personalized chemoprevention in early stage NSCLC patients. Supported by DoD grant W81XWH-10-1-1007.

Citation Format: {Authors}. {Abstract title} [abstract]. In: Proceedings of the 103rd Annual Meeting of the American Association for Cancer Research; 2012 Mar 31-Apr 4; Chicago, IL. Philadelphia (PA): AACR; Cancer Res 2012;72(8 Suppl):Abstract nr 1721. doi:1538-7445.AM2012-1721

Cancer Research
cancerres.aacrjournals.org

doi: 10.1158/1538-7445.AM2013-2367
Cancer Res April 15, 2013 73; 2367

Abstract 2367: Transcriptomic architecture of the airway field cancerization in early-stage non-small cell lung cancer .

Yuho Maki[1], Junya Fujimoto[1], Suk-Young Yoo[1], Adam Gower[2], Li Shen[1], Melinda M. Garcia[1], Mohamed Kabbout[1], Chi-Wan Chow[1], Waun Ki Hong[1], Neda Kalhor[1], Jing Wang[1], Cesar Moran[1], Avrum Spira[2], Kevin R. Coombes[1], Ignacio I. Wistuba[1], and Humam Kadara[1]

- Author Affiliations

[1]University of Texas MD Anderson Cancer Center, Houston, TX;
[2]Boston University, Boston, MA.

Proceedings: AACR 104th Annual Meeting 2013; Apr 6-10, 2013; Washington, DC

Abstract

Earlier work has identified in lung cancer a field cancerization (FC) phenomenon in which tumors and adjacent normal appearing tissues share specific molecular abnormalities (e.g., loss of heterozygosity) that may be highly pertinent to cancer pathogenesis. We sought to characterize the global molecular airway FC adjacent to early-stage non-small cell lung cancer (NSCLC) in an attempt to unravel profiles that may help to explain the development of the disease. We performed whole-transcript expression profiling of a set of resected early-stage NSCLC specimens (n=20 patients) with matched histologically normal airways of varying distance from the tumor and paired uninvolved normal lung tissue (n=194 samples). Using linear mixed-effects models, we derived FC profiles signifying genes concordantly differentially expressed between tumors and airways compared to normal lung tissues. Gene set enrichment analysis demonstrated that a subset of the genes (n=299) was significantly and congruently modulated between large airways of smokers with and without lung cancer. We then questioned whether the airway FC exhibits site from tumor-dependent expression patterns.

Ordinal regression analysis identified airway profiles (n=422 genes) that were significantly progressively expressed by distance from tumors and topologically organized into canonical cancer-associated pathways, such as eukaryotic initiation factor, p70S6K kinase, polo-like kinase and mammalian target of rapamycin signaling (all p<0.001). In addition, the site-dependent airway profiles recapitulated NSCLC expression patterns and were concordantly modulated between tumors and uninvolved normal lung tissues pinpointing their probable roles in lung cancer pathogenesis. Quantitative real-time PCR (QRTPCR) analysis confirmed the differential expression of FC markers selected by both pathways analysis and statistical criteria. Notably, lysosome associated protein transmembrane 4 beta (LAPTM4B), a putative oncogene with no known role in lung carcinogenesis, was among the top 5 site-dependent FC markers and was significantly elevated in NSCLC and immortalized bronchial epithelial cell lines compared to normal cells. Furthermore, transient or stable knockdown of LAPTM4B by RNA interference decreased

NSCLC cell growth as well as anchorage-dependent and -independent colony formation. In conclusion, our efforts in understanding the adjacent molecular

FC in NSCLC unraveled airway profiles that 1) are, in part, relevant to lung cancer detection; 2) are modulated by distance from corresponding tumors; 3) recapitulate NSCLC expression patterns and 4) harbor markers engaged in mediating the lung malignant phenotype. Profiling the adjacent airway FC in conjunction with tumors, may provide additional insights into the molecular pathology of NSCLC. Funded in part by Department of Defense award W81XWH-10-1-1007.

Citation Format: Yuho Maki, Junya Fujimoto, Suk-Young Yoo, Adam Gower, Li Shen, Melinda M. Garcia, Mohamed Kabbout, Chi-Wan Chow, Waun Ki Hong, Neda Kalhor, Jing Wang, Cesar Moran, Avrum Spira, Kevin R. Coombes, Ignacio I. Wistuba, Humam Kadara. Transcriptomic architecture of the airway field cancerization in early-stage non-small cell lung cancer . [abstract]. In: Proceedings of the 104th Annual Meeting of the American Association for Cancer Research; 2013 Apr 6-10; Washington, DC. Philadelphia (PA): AACR; Cancer Res 2013;73(8 Suppl):Abstract nr 2367. doi:10.1158/1538-

Cancer Research
cancerres.aacrjournals.org

doi: 10.1158/1538-7445.AM2014-1597
Cancer Res October 1, 2014 74; 1597

Abstract 1597: Transcriptomic architecture of the field of cancerization in the adjacent normal-appearing airway: Early mechanisms in lung carcinogenesis

Yuho Maki[1], Junya Fujimoto[1], Suk-Young Yoo[1], Melinda Garcia[1], Adam Gower[2], Li Shen[1], Chi-Wan Chow[1], Carmen Behrens[1], Neda Kalhor[1], Cesar Moran[1], Jing Wang[1], Avrum Spira[2], Kevin R. Coombes[3], Ignacio I. Wistuba[1], and Humam Kadara[1]

Author Affiliations

[1]UT MD Anderson Cancer Center, Houston, TX;
[2]Boston University, Boston, MA;
[3]Ohio State University, Columbus, OH.

Proceedings: AACR Annual Meeting 2014; April 5-9, 2014; San Diego, CA

Abstract

Increasing our understanding of early events in the pathogenesis of lung cancer is crucial for identification of new targets for prevention and treatment of this malignancy. Earlier work has shown that seemingly normal cells adjacent to the tumor carry specific molecular alterations that are characteristic of the tumor itself suggestive of a field of cancerization. By sampling and studying normal-appearing tissue, the molecular field of cancerization provides biological insights into early phases in cancer development. In this study, we sought to characterize molecular field effects in the normal-appearing airway that are most representative of the nearby lung tumor, and thus, are most likely to denote early events in lung carcinogenesis. To achieve this, we performed genome-wide expression profiling of resected field cancerization specimens (n=20 patients) comprised of matched early-stage non-small cell lung cancers (NSCLCs), cytologically normal airways with varying spatial distance from the tumors and distant (relative to location of tumors) normal lung tissues (n=194 samples). Using ordinal logistic regression, we identified 422 genes that were progressively modulated in expression in normal-appearing airways by spatial distance from tumors. Notably, when examined in paired NSCLC and normal lung tissues, these genes were found to recapitulate tumor expression profiles. We then sought to examine the role of lysosomal protein transmembrane 4 beta (LAPTM4B), a putative oncogene that was found to be up-regulated in airways by shorter spatial distance from tumors, in lung oncogenesis. LAPTM4B was significantly elevated in NSCLC tissues compared to paired distant normal lung and was predictive of poor survival in lung adenocarcinoma. Moreover, LAPTM4B promoted anchorage-dependent and -independent lung cancer cell growth and was crucial for cellular survival and the autophagy response under nutrient- and serum-deprived conditions. In addition, pathways analysis of a LAPTM4B-dependent gene expression profile revealed decreased activation of the canonical nuclear factor erythroid 2-like 2 (NRF2)-mediated pathway following LAPTM4B knockdown. Further, we found that LAPTM4B augmented the expression and nuclear translocation of the NRF2 transcription factor following serum deprivation pointing to the probable role of the novel LAPTM4B/NRF2 signaling axis in promoting lung cancer cell survival. All in all, our study points to molecular field of cancerization profiles in the normal-appearing airway that highly signify the nearby lung tumor and comprise early mechanisms (e.g. LAPTM4B) in lung carcinogenesis.

Citation Format: Yuho Maki, Junya Fujimoto, Suk-Young Yoo, Melinda Garcia, Adam Gower, Li Shen, Chi-Wan Chow, Carmen Behrens, Neda Kalhor, Cesar Moran, Jing Wang, Avrum Spira, Kevin R. Coombes, Ignacio I. Wistuba, Humam Kadara. Transcriptomic architecture of the field of cancerization in the adjacent normal-appearing airway: Early mechanisms in lung carcinogenesis. [abstract]. In: Proceedings of the 105th Annual Meeting of the American Association for Cancer Research; 2014 Apr 5-9; San Diego, CA. Philadelphia (PA): AACR; Cancer Res 2014;74(19 Suppl):Abstract nr 1597. doi:10.1158/1538-7445.AM2014-1597

Cancer Research
cancerres.aacrjournals.org

doi: 10.1158/1538-7445.AM2014-2352
Cancer Res October 1, 2014 74; 2352

Abstract 2352: Mapping the airway-wide molecular field of injury in smokers with lung cancer

Rebecca Kusko[1], **Christina Anderlind**[2], **Gerald Wang**[3], **Sherry Zhang**[1], **W. Dean Wallace**[3], **Tonya Wasler**[3], **Michael Ebright**[1], **Melinda M. Garcia**[4], **Rosana Eisenberg**[5], **Gina Lee**[3], **Gang Liu**[1], **David Elashoff**[3], **Neda Kalhor**[4], **Cesar Moran**[4], **Reza Mehran**[4], **Junya Fujimoto**[4], **Pierre P. Massion**[5], **Steven Dubinett**[3], **Ignacio Wistuba**[4], **Marc Lenburg**[1], **Humam Kadara**[4], and **Avrum Spira**[1]

Author Affiliations

[1]Boston University School of Medicine, Boston, MA;
[2]Tufts Medical Center, Boston, MA;
[3]David Geffen School of Medicine at UCLA, Los Angeles, CA;
[4]The University of Texas MD Anderson Cancer Center, Houston, TX;
[5]Thoracic Program, Vanderbilt Ingram Cancer Center, Nashville, TN.

Proceedings: AACR Annual Meeting 2014; April 5-9, 2014; San Diego, CA

Abstract

Lung cancer mortality is the leading cause of cancer death in the United States in part because diagnosis occurs after regional or distant metastasis of the disease. Identifying effective early detection biomarkers is crucial for improving lung cancer clinical management. Moreover, molecular biomarkers for early disease detection may provide insight into the molecular pathways associated with disease development and progression. Our lab has shown that smoking-induced gene expression alterations are mirrored in the epithelia of the mainstem bronchus, buccal and nasal cavity. We have additionally demonstrated that gene-expression profiles in cytologically normal mainstem bronchial epithelium can serve as an early diagnostic biomarker for lung cancer. Here we expand on our previous work by spatially mapping the molecular field of injury throughout the entire respiratory tract in smokers with lung cancer. Using Affymetrix Gene ST 2.0 arrays, we profiled genome-wide gene-expression in 1) lung lesions and adjacent normal lung obtained from smokers undergoing surgical resection, 2) epithelial brushings obtained at intraoperative bronchoscopy from the nasal epithelium, main carina and ipsilateral and contralateral proximal and distal bronchi (relative to the location of the resected lung lesion), and 3) epithelial brushings obtained at lobectomy from sub-segmental bronchus (adjacent to tumor). Linear modeling approaches comparing the airways and tumors of patients with cancer to those with benign lung disease were used to explore relationships in cancer-specific gene-expression alterations across sites within the respiratory tract. We found that genes upregulated in the small airways leading to the tumor were enriched in genes upregulated in the mainstem bronchus and main carina of smokers with lung cancer. In addition, genes upregulated in the bronchus and main carina of smokers with lung cancer showed enrichment among cancer associated genes elevated in the nose. Furthermore, a linear mixed effects model uncovered genes and pathways which change in expression in a gradient-like manner as distance from the tumor increases. Our findings suggest that the molecular field of injury encompasses airway-wide alterations throughout the entire respiratory tract of smokers with lung cancer as well as gradient profiles that change with respect to proximity of the nearby tumor. These molecular alterations may ultimately serve as early detection biomarkers for lung cancer and provide new insights into early stages of lung carcinogenesis.

Citation Format: Rebecca Kusko, Christina Anderlind, Gerald Wang, Sherry Zhang, W. Dean Wallace, Tonya Wasler, Michael Ebright, Melinda M. Garcia, Rosana Eisenberg, Gina Lee, Gang Liu, David Elashoff, Neda Kalhor, Cesar Moran, Reza Mehran, Junya Fujimoto, Pierre P. Massion, Steven Dubinett, Ignacio Wistuba, Marc Lenburg, Humam Kadara, Avrum Spira. Mapping the airway-wide molecular field of injury in smokers with lung cancer. [abstract]. In: Proceedings of the 105th Annual Meeting of the American Association for Cancer Research; 2014 Apr 5-9; San Diego, CA. Philadelphia (PA): AACR; Cancer Res 2014;74(19 Suppl):Abstract nr 2352. doi:10.1158/1538-7445.AM2014-2352

Official Journal of the Asian Pacific Society of Respirology
Respirology

INVITED REVIEW SERIES:
TRANSLATING RESEARCH INTO PRACTICE
SERIES EDITOR: JOHN E HEFFNER AND DAVID CL LAM

Pulmonary adenocarcinoma: A renewed entity in 2011

HUMAM KADARA,[1] MOHAMED KABBOUT[1] AND IGNACIO I. WISTUBA[1,2]

Departments of [1]Thoracic/Head and Neck Medical Oncology and [2]Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

ABSTRACT

Lung cancer, of which non-small-cell lung cancer comprises the majority, is the leading cause of cancer-related deaths in the United States and worldwide. Lung adenocarcinomas are a major subtype of non-small-cell lung cancers, are increasing in incidence globally in both males and females and in smokers and non-smokers, and are the cause for almost 50% of deaths attributable to lung cancer. Lung adenocarcinoma is a tumour with complex biology that we have recently started to understand with the advent of various histological, transcriptomic, genomic and proteomic technologies. However, the histological and molecular pathogenesis of this malignancy is still largely unknown. This review will describe advances in the molecular pathology of lung adenocarcinoma with emphasis on genomics and DNA alterations of this disease. Moreover, the review will discuss recognized lung adenocarcinoma preneoplastic lesions and current concepts of the early pathogenesis and progression of the disease. We will also portray the field cancerization phenomenon and lineage-specific oncogene expression pattern in lung cancer and how both remerging concepts can be exploited to increase our understanding of lung adenocarcinoma pathogenesis for subsequent development of biomarkers for early detection of adenocarcinomas and possibly personalized prevention.

Key words: airway epithelium, field cancerization, genetics, lung cancer, molecular biology.

INTRODUCTION

Lung cancer is the leading cause of cancer deaths in the United States and worldwide in both developing and developed regions.[1] The high mortality of this disease is in part due to the late diagnosis of the majority of lung cancers after regional or distant spread of the malignancy[2] and when only palliative treatment options are available.[3] Given that various epithelial tumours develop in a multi-stage stepwise fashion, it is plausible to assume that early diagnosis of lung cancer or intraepithelial lesions coupled with effective prevention strategies will improve survival of patients and reduce the significant health burden and mortality associated with this disease.[3] Despite recent encouraging findings from the National Lung Screening Trial (NLST),[4] early detection of lung cancer is challenging due to the lack of biomarkers for early diagnosis of the disease and to the presence of multiple neoplastic molecular pathways that mediate lung carcinogenesis. A better understanding of the molecular origins of lung cancer is expected to pave the way for unmet effective and personalized strategies for lung cancer prevention and treatment.

The two major forms of lung cancer are non-small-cell lung cancer (NSCLC), which accounts for approximately 85% of all diagnosed lung cancers, and small-cell lung cancer (SCLC), which constitute about 15% of lung neoplasms.[2] NSCLC is comprised of three major histological subtypes, squamous-cell carcinomas (SCC), lung adenocarcinomas and large-cell lung carcinomas.[2,5] Several major differences exist between adenocarcinomas and SCC, the two major subtypes of NSCLC. Compared with SCC and SCLC that arise from the major bronchi and are centrally located, pulmonary adenocarcinomas arise from small bronchi, bronchioles or alveolar epithelial cells, and are typically peripherally located as reviewed elsewhere.[2,5–7]

The Authors: Humam Kadara, PhD, is an instructor in the Department of Thoracic/Head and Neck Medical Oncology, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center. His research interests focus on lung cancer genomics, pathogenesis and prevention. Mohamed Kabbout, PhD, is a postdoctoral fellow in the same department researching on mutant Kirsten rat sarcoma oncogene-mediated lung cancer pathogenesis. Ignacio I. Wistuba, MD, is a Jeri and Lori Eisenberg Professor of Pathology in the Departments of Thoracic/Head and Neck Medical Oncology and Pathology and director of the Thoracic Molecular Pathology Laboratory at the University of Texas MD Anderson Cancer Center. His research interests focus on understanding the molecular pathology of lung cancer to guide or develop therapeutic and prevention strategies.

Correspondence: Humam Kadara, Department of Thoracic/Head and Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, Email: hkadara@mdanderson.org

Received 10 October 2011; accepted: 11 October 2011.

doi: 10.1111/j.1440-1843.2011.02095.x

Clinically, SCC and lung adenocarcinoma respond differently to chemotherapeutic agents, exemplified by the use of pemetrexed for treatment of the latter subtype and not for SCC.[8,9] Moreover, although smoking is the major causative factor in lung cancer pathogenesis, significant differences in smoking patterns are observed between the two major NSCLC histological subtypes. Whereas SCC pathogenesis is strongly linked to smoking, lung adenocarcinoma is the more common histological subtype in never-smoker patients.[10–13] Accumulating evidence suggests that lung adenocarcinoma arising in never-smokers is a disease with different pathological and epidemiological features compared with adenocarcinomas causally linked to cigarette smoking.[13] Specifically, never-smoker lung adenocarcinoma is more commonly diagnosed in females compared with males[14] and is more frequently found in eastern and southern parts of the Asian continent,[15] and displays better prognosis and survival compared with ever-smoker patients.[2,12,13] At the molecular level, and to date, two major pathways are thought to mediate lung adenocarcinoma development: an epidermal growth factor receptor (*EGFR*)-dependent pathway in never-smokers and a Kirsten rat sarcoma oncogene (*KRAS*)-dependent signalling module in smokers[16–23] (discussed further later). Further understanding of lung adenocarcinoma pathogenesis would be needed to unravel other pathways that play important roles in development of this major subtype of lung cancer.

Lung adenocarcinomas have a wide spectrum of clinical, molecular and histological features.[24] The 2004 World Health Organization (WHO) classification of lung tumours included four growth patterns for its adenocarcinoma classification: bronchioloalveolar (BAC; also known as lepidic), acinar, papillary and solid.[24] Most invasive lung adenocarcinomas are heterogeneous in nature and include more than one of these histological patterns.[24,25] The existing lung adenocarcinoma histological heterogeneity and the varying clinicopathological features (e.g. patient outcome) of the aforementioned histological patterns highlight the importance of incorporating histological pattern information into clinical management of this complex disease. More recently, the European Respiratory Society (ERS), the International Association for the Study of Lung Cancer (IASLC) and the American Thoracic Society (ATS) sponsored a new classification of lung adenocarcinoma.[26] The new classification study presented several modifications to the WHO 2004 criteria for diagnosis of resected adenocarcinoma specimens. Mainly, the consortium study suggested that the term BAC should be discontinued.[26] Instead, it is agreed that adenocarcinoma *in situ* (AIS) and minimally invasive adenocarcinoma (MIA) are to be used for small adenocarcinomas with either pure lepidic growth or predominant lepidic growth with less than 5 mm invasion, respectively.[26] Moreover, the new classification dropped the use of mixed subtype, and instead, adenocarcinomas are classified according to their predominant subtype.[26]

This review will describe advances in the molecular pathology of lung adenocarcinoma with emphasis on genomics and DNA alterations of this disease. Moreover, the review will describe recognized lung adenocarcinoma preneoplastic lesions and current concepts of the early pathogenesis and progression of the disease. We will also portray the field cancerization phenomenon and lineage-specific oncogene expression pattern in lung cancer and how both remerging concepts can be exploited to increase our understanding of lung adenocarcinoma pathogenesis for subsequent development of biomarkers for early detection of adenocarcinomas and possibly personalized prevention.

REVIEW

Molecular pathology of lung adenocarcinoma

Lung adenocarcinomas exhibit unique genomic aberrations compared with lung SCC, indicating that the molecular pathology of both NSCLC subtypes encompasses different molecular pathways of development and progression.[2] Earlier studies have shown that lung SCC exhibit higher frequencies of deletions at chromosomal regions 17p13 (*TP53*), 13q14 (*RB*), 9p21 (*CDKN2A*), 8p21–23 and 3p compared with lung adenocarcinomas.[27–29] Moreover, many of the aforementioned molecular abnormalities (e.g. allelic losses at 9p21 and 13q24) occur in the sequential multi-step progression of SCC but not of adenocarcinomas.[6,27] In contrast, mutations in the *KRAS*, *EGFR* and *HER2/NEU* oncogenes occur almost exclusively in adenocarcinomas.[2,20,22,23,30,31] Amplification of the embryonic stem cell (ESC) factor sex determining Y-box 2 (*SOX2*) is exclusive to SCC,[32,33] and increased gene dosage and protein expression of thyroid transcriptional factor–1/NK2 homeobox 1 (*TITF-1/NKX2-1*) is prevalent in lung adenocarcinoma, indicating that both transcriptional factors most likely function as lineage-specific genes in lung cancer.[34–36] This section will highlight molecular abnormalities, with special emphasis on genomics and DNA alterations, of lung adenocarcinoma that render this malignancy a unique entity.

Lung adenocarcinoma genomics

KRAS, a low molecular weight guanosine triphosphatase (GTPase) and the major upstream activator of the RAF-MEK-ERK pathway, is considered to be the most frequently mutated oncogene in lung adenocarcinomas.[19,21,37] As mentioned before, mutations in this oncogene are more common in adenocarcinomas arising in ever-smoker (former and current) lung cancer patients.[13,16,17,19,20,22,23,30] Most *KRAS* mutations involve replacing glycine 12 with other amino acids such as valine (G12V), aspartic acid (G12D) and glutamic acid (G12D), and replacing glycine 13, and are activating rendering the gene with reduced GTPase activity with subsequent potent activation of mitogenic and proliferative signalling through the RAF-MEK-ERK cascade.[19,37–39] Thus, it is plausible to assume that therapeutic strategies targeting *KRAS*

would be very beneficial in adenocarcinomas with activating mutations in this oncogene. However, there are currently no available treatment options for *KRAS*-mutant lung adenocarcinomas compared with tumours with mutations in other oncogenes,[40] as strategies targeting *KRAS* farnesylation, *MEK* activation and *BRAF* have either failed or yielded no responses.[41–43]

In contrast to *KRAS*, mutations in *EGFR* are strongly linked to lung adenocarcinomas arising in never-smokers and are suggested to molecularly drive the disease in this patient subpopulation.[13,14,17,18,22,23,30,37] It is important to note that *EGFR* mutations are more common in East Asian patients and in female gender.[2,13,22] Small in-frame deletions in exon 19 and missense mutations in exon 21 (L858R and L861Q) are the most common mutations detected in *EGFR*[44] and were shown by several ground-breaking studies to underlie sensitivity of lung adenocarcinoma patients to *EGFR*-targeting small tyrosine-kinase inhibitors (e.g. erlotinib and gefitinib).[18,45,46] These studies were the first to prove the feasibility of personalized medicine approaches for the management of lung adenocarcinoma and represent the landmark for the application of genomic medicine in this disease.

The discovery of fusions involving anaplastic large-cell lymphoma kinase (*ALK*) with the upstream partner echinoderm microtubule associated protein 4 (*EML4*) by Soda *et al.*[47] further opened new venues for genomic-driven personalized treatment strategies for lung adenocarcinoma.[48] Both *EML4* and *ALK* are located in chromosome 2p, and fusion of both involves small inversions within this region.[47] *EML4-ALK* fusion results in constitutive activation of the *ALK* kinase rendering cells and adenocarcinoma tumours expressing this oncogenic fusion protein sensitive to *ALK* inhibitors.[47–49] Like *EGFR* mutations, *EML4-ALK* fusion genes are prevalent in lung adenocarcinomas, younger patients and, in particular, in lifetime never-smoker patients or light smokers.[49,50] Importantly, *EML4-ALK* fusion genes are mutually exclusive from *EGFR* and *KRAS* mutations, indicating that such molecular defects function as drivers of pathogenesis, which is clinically important, as it increases potential of personalized treatment options that target driver oncogenes in this malignancy.[50]

Other mutually exclusive and, thus, potential oncogenic drivers have been identified in lung adenocarcinomas. Mutations in *HER2/NEU* were found by Stephens *et al.* to occur in lung adenocarcinomas.[31] Compared with mutations of *EGFR* oncogene, *HER2/NEU* mutations are less frequent[30] and have not been successfully exploited in the clinic for lung adenocarcinoma treatment.[51] Similar to *HER2/NEU*, mutations in *BRAF* also occur at low frequency in lung adenocarcinoma and are exclusive from *EGFR* and *KRAS* mutations, as well as from *EML-ALK* fusions.[50] There are yet no successful target-specific treatment strategies for lung adenocarcinoma with *BRAF* mutations. It is important to note that mutations in *HER2/NEU* and *BRAF* have not been found in lung SCC.[50]

The *TP53* tumour suppressor is the most frequently mutated gene in lung adenocarcinoma (65–70%). Various abnormalities in *TP53* were identified in lung adenocarcinoma almost two decades ago[52,53] and more recently in the tumour-sequencing project[37] and occur in similar pathways to those mediated by the oncogenic driver mutations mentioned earlier.[20,22,37] Mutations in the *CDKN2A* tumour suppressor have also been described in lung adenocarcinoma.[54,55] However, methylation[56] or focal DNA deletion[36,55] rather than mutation of this tumour suppressor seems to be more frequent and occurs earlier in lung cancer pathogenesis.[5] With the advent of various technologies including single nucleotide polymorphism (SNP) arrays, mass spectrometry mutational analysis and more recently second-generation sequencing, and the undertaking of large-scale studies such as the tumour-sequencing project,[37] our knowledge of the mutational spectrum of lung adenocarcinoma has substantially increased. New mutated oncogenes and tumour suppressor genes have been identified in lung adenocarcinoma and along with previously characterized mutated genes are outlined in Table 1 and have been reviewed in detail elsewhere.[50,51] It is important to note that many, if not

Table 1 Mutations in lung adenocarcinoma

—	Mutation rate (%)
Oncogenes	
KRAS	15–30
EGFR	5–40
ALK (fusion)	5–15
MET	14
KDR	5
EPHA3	5
MAP2K1	5
HER2/ERBB2	2–4
FGFR4	4
PDGFRA	4
NTRK1	4
NTRK3	4
EPHA5	4
ERBB4	4
LTK	3
PAK3	3
ERBB3	2
FGFR1	2
FGFR2	2
NRAS	2
PIK3CA	2
BRAF	2
AKT1	1
Tumour-suppressor genes	
TP53	50–70
STK1	20–30
LRP1B	9
NF1	7
ATM	7
APC	6
PTPRD	5
CDKN2A	5
RB1	4
PTEN	2

most, of these mutations are not mutually exclusive of other driver mutations and events such as *EGFR* mutations and *EML4-ALK* fusions. For example, *PIK3CA* mutations were always found together with *EGFR* mutations in never-smoker lung adenocarcinomas.[57] It is also worthwhile to mention that some of the outlined mutations have been detected in both lung adenocarcinomas and SCC (e.g. *PIK3CA* and *MET*) or only in the former subtype of NSCLC (e.g. *MEK1*, *HER2/NEU* and *BRAF*).[50] It is unknown for most of mutations occurring in lung adenocarcinomas and recently identified by exon-directing sequencing in the tumour-sequencing project, whether they also occur in lung SCC. The discovery of new oncogene and tumour suppressor mutations in lung adenocarcinoma occurring in mutually exclusive and inclusive cell signalling pathways expands the range of possible target-specific and even combinatorial personalized therapeutic strategies for this disease.

Copy-number alterations

Gene dosage variations occur in many pathological conditions. For example, in cancer, deletions and copy-number increases modulate the expression of tumour-suppressor genes and oncogenes, therefore contributing to tumourigenesis. Characterization of these DNA copy-number changes is vital for both the basic understanding of cancer and its diagnosis. Copy-number alterations are routinely assessed in laboratories by fluorescent *in situ* hybridization (FISH) techniques as well as genomic polymerase chain reaction (PCR), including quantitative PCR approaches. However, these approaches are labour-intensive and would hamper the discovery and complete understanding of the genome of tumours in large-scale studies. High-throughput and genome-wide analysis of DNA copy-number alterations was made possible by comparative genomic hybridization (CGH) approaches, which utilize differentially labelled test and reference genomic DNA that are co-hybridized to normal metaphase chromosomes.[58] CGH, however, exhibits limited mapping resolution even when compared with lower throughput higher resolution techniques, such as FISH.[59] Subsequently, high-resolution genome-wide analysis was successfully performed using cDNA microarray-based CGH and SNP arrays coupled with statistical methods to assess both the amplitude and the frequency of copy-number changes at each position in the genome.[59]

Genome-wide alterations in human lung adenocarcinoma tumours stemming from several major studies have assuredly increased our understanding of the molecular pathogenesis of this major malignancy.[36,60,61] Earlier chromosomal CGH studies have revealed in NSCLC recurrent gains at 1q31, 3q25–27, 5p13–14 and 8q23–24, and deletions at 3p21, 8p22, 9p21–22, 13q22 and 17p12–13.[62–66] Moreover, these early studies already highlighted genomic differences and similarities between lung adenocarcinomas and SCC; most prominent of which were gains in 3q mainly by lung SCC.[62–66] For example, Petersen *et al*. found genomic aberrations that characterize lung adenocarcinomas from SCC, namely gain of 1q23, and the deletion at 9q22 were significantly associated with adenocarcinomas, whereas the loss of chromosomal band 2q36–37 and gain of 3q were strongly associated with SCC.[66] Bjorkqvist *et al*. demonstrated that 94% (15/16) of lung SCC analysed had a gain in 3q, whereas only 24% (4/17) of the adenocarcinoma samples exhibited a gain in 3q, and high-level amplifications in 3q were only detected in SCC.[63] In addition, Luk *et al*. demonstrated that gains at 1q22–32.2, 15q, 20q and losses at 6q, 13q and 18q were more prevalent in lung adenocarcinomas, whereas SCC, as shown in earlier studies, exhibited gains/amplifications at 3q.[64] Moreover, Pei *et al*. showed that besides prevalent gain of 3q in lung SCC, gain of 20p13 and loss of 4q also were significantly higher in SCC, whereas gain of 6p was more common in adenocarcinomas.[65] Massion *et al*. utilized higher resolution analysis by array CGH to study copy-number alterations of known loci and found that the most distinct genomic aberration between both NSCLC subtypes was gain of 3q22–26 and loss of 3p by lung SCC.[67] Moreover, and in the same study, *PIK3CA* oncogene was found to be a member of the chromosome-3q amplicon with higher copy number and expression in SCC but not in adenocarcinomas.[67]

More recent studies utilized more advanced technologies to query the genome of lung adenocarcinoma and associate specific gene modulations with chromosomal and loci gain or losses. In the study by Tonon *et al*., high-resolution cDNA microarray-based CGH was utilized to study the genomic profiles of 18 lung adenocarcinomas and 26 SCC, as well as 14 NSCLC cell lines.[61] The study identified 93 focal copy-number alterations that mainly comprised previously uncharacterized recurrent high-amplitude amplifications and homozygous deletions.[61] Besides confirming previous findings by chromosomal CGH and highlighting known gains (e.g. 1q31 and 3q25–27) and known deletions (e.g. 3p, 8p22 and 13q22), the study by Tonon *et al*. was able to map specific genes to focal copy-number alterations including *CDKN2A* and *RB1* tumour-suppressor genes and *EGFR*, and *KRAS* oncogenes.[61] However, when comparing both adenocarcinomas and lung SCC, the study found that the only notable genomic difference between both NSCLC histological subtypes was the well-characterized gain of 3q26–29 in SCC that included *TP63*, well known for its role in squamous differentiation, and concluded that similar oncogene and tumour-suppressor gene aberrations drive lung adenocarcinoma and SCC pathogenesis.[61] The study by Tonon *et al*. was a major step in understanding the genomic profiles of NSCLC tumours despite the small number of lung adenocarcinomas and SCC analysed.

Later on, Weir *et al*. studied the genomic profiles of a large collection of primary lung adenocarcinomas ($n = 371$) by high-density SNP arrays using 238 000 probe sets.[36] The report by Weir *et al*. was a milestone in understanding the lung adenocarcinoma genome, as it unravelled previously uncharacterized amplified genes and loci that otherwise may have not been

identified using a small number of primary tumours. The study identified 39 large-scale chromosomal arm gain or loss, 26 of which were significantly recurrent across many lung adenocarcinomas. Importantly, the large-scale study by Weir *et al*. identified 31 recurrent focal events that included 24 amplifications and 7 homozygous deletions. Using dense SNP arrays coupled with statistical methods (genomic identification of significant targets in cancer), the group was able to associate specific genes to the focal events and rank significance of events based on both the amplitude and frequency of copy-number change,[36] similar to what was performed by Tonon *et al*.[61] to identify minimal common regions of copy-number alterations. The most significant focal regions of amplification included known oncogenes such as *MDM2* (12q15), *MYC* (8q24), *EGFR* (7p11), *CDK4* (12q14), *KRAS* (12p12), *CCNE1* (19q12), *ERBB2* (17q12), *CCND1* (11q13) and *TERT* (5p15).[36] It is important to note that three of these oncogenes, *EGFR*, *KRAS* and *ERBB2*, are mutated in lung adenocarcinoma, as discussed before, suggesting that amplification and mutation of these oncogenes may cooperate systematically in lung adenocarcinoma pathogenesis. The most significant focal regions of deletions also included known tumour-suppressor genes such as *CDKN2A* (9p21) and *PTEN* (10q23).[36] Although 5p was previously shown to be gained in lung adenocarcinomas, the identity of genes involved in this gain was unknown prior to the study by Weir *et al*. The application of more advanced technologies to characterize the genomic profile of lung adenocarcinomas enabled the group to highlight previously unknown associations between canonical cancer-associated genes and known loci copy-number alterations, as well as to identify potentially new oncogenes. For example, 10 genes, including *TERT*, were found in the study to be included in the 5p15 region.[36] Furthermore, the study highlighted previously uncharacterized amplification of *TITF-1/NKX2-1* (14q13.3) in lung adenocarcinomas and demonstrated the oncogenic role of this lineage-specific transcriptional factor in lung cancer cells evidenced by the effect of RNA interference-mediated knockdown of its expression on anchorage-independent growth of lung adenocarcinoma cell lines with amplification of this gene.[36] The amplification/copy-number gain of *TITF-1* in lung adenocarcinoma was later confirmed in different studies including that by Kwei *et al*. using array CGH.[34] However, it is important to mention that although *TITF-1* amplification is generally prevalent in lung adenocarcinoma and is thought to function as a lineage-specific oncogene in this subtype of NSCLC, it has also shown by FISH analysis in lung SCC, which will be discussed later in this review.

As mentioned before, numerous earlier studies have demonstrated that gain of 3q is a genomic feature of lung SCC. Similar to the aforementioned study by Weir *et al*., Bass and colleagues utilized high-density Affymetrix SNP arrays to analyse 40 oesophageal and 47 lung SCC, which confirmed that gain of 3q26 was the main focal amplification event in lung and oesophageal SCC.[32] Importantly, the same study revealed the presence of the ESC factor in this region, which was later on confirmed to be amplified in SCC and not in lung adenocarcinomas, and promoted the survival of cell lines of the former NSCLC subtype harbouring amplification of this transcriptional factor and further suggesting that lung adenocarcinomas are genetically different from SCC.

More recently, high-resolution array CGH was performed on never-smoker lung adenocarcinomas ($n = 60$) with known mutation status of *EGFR*.[68] This study identified 14 new minimal common regions of gain or loss and confirmed previously known copy-number alterations such as those involving *TERT* (5p), *TITF-1* (14q13), *EGFR* (7p) and *CDKN2B* (9p). Notably, the study revealed new genomic aberrations, namely the 16p11.2 region harbouring the *FUS* oncogene that functions in transcriptional splicing and DNA repair.[68] Gain of 16p11.2 was evident in greater than 20% of the never-smoker lung adenocarcinomas analysed and mRNA levels of *FUS* correlated with copy gain of 16p, as they were higher in tumours with gain of this region compared with tumours that did not exhibit 16p copy gain. Importantly, the study by Job *et al*. revealed genomic copy-number alterations that were highly associated with presence of *EGFR* mutation, an oncogenic driver of never-smoker lung adenocarcinoma pathogenesis.[68] Gains of 7p were significantly associated with presence of *EGFR* mutations and included *EGFR* gene, suggesting, as mentioned before, that copy-number alternations and mutations cooperate at the genomic level in lung adenocarcinoma pathogenesis. However, it cannot be ignored that *EGFR* copy gain or amplification may favour the detection of *EGFR* mutations in a heterogeneous tumour due to the mutant allele-specific imbalance phenomenon.[69] In a more recent study by Yuan *et al*. and also using array CGH technology, gains in 7p, including the *EGFR* gene, were common in *EGFR* mutant lung adenocarcinomas and predicted overall and recurrence-free survival in this disease population.[70] More importantly, in contrast to *EGFR* mutations, presence of genes (including *EGFR*) within the 7p gain predicted poorer response to tyrosine kinase inhibitors targeting *EGFR*.[70]

Gene expression profiling

Numerous studies have utilized microarray technology to analyse the global transcriptome of NSCLC for diagnosis (discussed later), molecular classification, response to therapy and prognosis. For the purpose of this review, we will discuss several key studies that investigated expression profiles of lung adenocarcinomas to further understand the molecular biology of this prevalent lung malignancy. Bhattacharjee and colleagues utilized arrays to study adenocarcinomas of lung origin ($n = 127$), SCC ($n = 21$), carcinoids ($n = 20$), SCLC ($n = 6$) and 17 normal samples.[71] The study found that differential expression profiles segregated the samples into different clusters based on histology, evidenced by the two-dimensional cluster

analysis. Genes associated with squamous differentiation such as keratin and *TP63* were overexpressed in SCC, and neuroendocrine markers were enriched in the SCLC cluster. Importantly, the study by Bhattacharjee *et al*. also analysed the adenocarcinomas alone by hierarchical clustering and demonstrated that the adenocarcinomas were heterogeneous in molecular make-up, being separated into various clusters with distinct clinical outcomes.[71] Similarly, Garber *et al*. utilized cDNA microarrays to study expression profiles of 41 adenocarcinomas, 16 SCC, 5 large-cell carcinomas and 5 SCLC, as well as 5 normal lung samples.[72] Again, lung adenocarcinomas were most heterogeneous and were divided into different clusters that were associated with clinicopathological variables such as tumour grade.[72] Later, Hayes *et al*. found that adenocarcinoma subtypes identified by the Bhattacharjee and Garber studies were reproducible in additional microarray datasets.[73] Several other studies have also demonstrated, using microarray expression profiling technology, the heterogeneity of lung adenocarcinomas and their distinction from other lung cancer subtypes. As reviewed by Yatabe, global gene expression profiling was able to subdivide lung adenocarcinomas into various clusters that correlated with *EGFR* mutation status, prognosis, expression of lung peripheral airway markers such as surfactant proteins (SP) and *CC10*, as well as enrichment of the BAC subtype.[44]

Lung adenocarcinoma preneoplasia

From biological and histopathological perspectives, NSCLC is a complex malignancy that develops through multiple preneoplastic pathways. Lung adenocarcinoma, a major subtype of NSCLC, has been increasing in incidence globally in both smokers and non-smokers[13] with a concurrent decrease in SCC frequency. It has been postulated that the increasing incidence of lung adenocarcinomas compared with SCC is in part due to the change in the type of cigarettes used (lower nicotine and tar) and smoking habits and behaviour.[11] Anatomical differences in the location of diagnosed lung adenocarcinomas and SCC strongly suggest that both NSCLC subtypes develop through different histopathological and molecular pathways and have different cells of origin; however, the specific respiratory epithelial cell type from which each lung cancer type develops has not been established with certainty.[5] Lung SCC is typically centrally located in the lung and is thought to arise from the major bronchi. In contrast, lung adenocarcinomas that are usually peripherally located are believed to arise from small bronchi, bronchioles or alveoli of the distant airways of the lung. The sequence of histopathological changes in bronchial epithelia that precede the development of lung SCC has been characterized.[6,27] However, the sequential preneoplastic changes, as well as the corresponding molecular abnormalities, leading to the development of lung adenocarcinomas are poorly documented.

Histopathological development of lung adenocarcinoma

Clara cells and the type II pneumocytes are believed to be the progenitor cells of the peripheral airways, and peripherally arising adenocarcinomas often express markers of these cell types.[44,74] Atypical adenomatous hyperplasias (AAH) are considered to be a precursor lesion for peripheral lung adenocarcinomas.[5,7] However, and until now, AAH is the only sequence of morphological change identified so far for the development of invasive lung adenocarcinomas, and there is consensus that the pathogenesis of many adenocarcinomas is largely unknown. The postulated progression of AAH to adenocarcinomas *in situ*, which is characterized by the growth of neoplastic cells along pre-existing alveolar structures without invasion, is supported by molecular studies.[75] Distinction between highly atypical AAH and what was known as BAC is sometimes difficult. Therefore, and as mentioned before, the ERS, IASLC and ATS sponsored a new classification of lung adenocarcinoma that presented several modifications to the WHO 2004 criteria for diagnosis of resected adenocarcinoma specimens. The term BAC was suggested to be discontinued and replaced with AIS and MIA used for small adenocarcinomas with either pure lepidic growth or predominant lepidic growth with less than 5 mm invasion, respectively. Importantly, the clinical features of both adenocarcinoma progression types are unique as patients with AIS or MIA have a 100% 5-year survival rate after respective surgery.[26]

The differentiation phenotype derived from immunohistochemical and ultrastructural features indicates that AAH originate from the progenitor cells of the peripheral airways.[26,44,74] Surfactant apoprotein and Clara cell-specific 10-kDd protein are expressed in almost all AAH. In addition, an increasing body of evidence suggests that AAH is the precursor of at least a subset of adenocarcinomas. For example, AAH is most frequently detected in lungs of patients bearing lung cancers (9–20%), especially adenocarcinomas (as many as 40%), compared with lung SCC (11%).[76] It is important to note that AAH is detected more frequently in East Asian patients relative to Western patients. In such studies, it has been suggested that AAH is involved in the linear progression of cells of the 'terminal respiratory unit' (TRU) to AIS and subsequently invasive adenocarcinomas[7,44,74] due to the expression of common genes between the TRU and the AAH, which is discussed later. Such studies have postulated that most, if not all, peripheral lung adenocarcinomas progress from alveoli through AAH as a preneoplastic lesion. As will be discussed further later, we have noted similar molecular abnormalities (e.g. *EGFR* mutations) between adenocarcinomas arising in never-smokers and small bronchioles within the localized and adjacent fields of the adenocarcinomas, suggesting that lung adenocarcinomas may arise from bronchiolar epithelium and small bronchi, and not only from alveoli.[77,78] In a recent review by Yatabe *et al*., a nonlinear progression schema for lung adenocarcinomas was suggested.[7] In this nonlinear schema, Yatabe *et al*. postulated that

lung adenocarcinomas of the TRU subtype, as named by the authors, develop through AAH. On the other hand, and according to the same nonlinear progression hypothesis, some lung adenocarcinomas arise through unknown preneoplastic precursors from other cells besides the TRU, which we believe may as well be the bronchiolar epithelium.[77,78]

Molecular pathogenesis of lung adenocarcinomas

Several molecular changes frequently present in lung adenocarcinomas are also present in AAH lesions, and they are further evidence that AAH may represent true preneoplastic lesions.[79] The most important finding is the presence of *KRAS* (codon 12) mutations in as many as 39% of AAH, which are also a relatively frequent alteration in lung adenocarcinomas.[6,80] Other molecular aberrations that were identified in AAH are overexpression of cyclin D1 (~70%), survivin (48%) and *HER2/neu* (7%) proteins.[5] Moreover, and as mentioned in the review by Wistuba and Gazdar, some AAH lesions were found to exhibit loss of heterozygosity (LOH) in chromosomes 3p (18%), 9p ($p16^{INK4a}$, 13%), 9q (53%), 17q and 17p (*TP53*, 6%).[5] It is noteworthy that most if not all of the aforementioned changes identified in AAH lesions are also frequently detected in lung adenocarcinomas. Later, AAH lesions were shown to exhibit LOH of tuberous sclerosis complex (TSC)-associated regions, activation of telomerase, loss of *LKB1*, overexpression of DICER, a key effector protein for small interfering RNA and miRNA function, and DNA methylation of *CDKN2A* and *PTPRN2*.[6,81,82] It is important to note that several studies have attempted to globally comprehend differential gene expression patterns and copy-number alterations between low-grade lesions (e.g. precursor lesions) or *in situ* adenocarcinomas and invasive tumours and found that amplification of the *EGFR* oncogene was the predominant differential molecular feature between the two different adenocarcinoma grade classes and occurred after mutations in the gene.[7] Importantly, as will be discussed in the next section of this review, *EGFR* mutations also preceded changes in copy number of the gene when studying histologically normal bronchiolar epithelia.[78]

KRAS and EGFR mutations in lung adenocarcinoma pathogenesis

Although there is only one sequence of morphological change characterized so far for the development of invasive lung adenocarcinomas, namely AAH, a large body of evidence suggests that at least two molecular pathways are involved, the *KRAS* and *EGFR* pathways in smoker and never-smoker adenocarcinoma subpopulations, respectively.[2,14,16,17,20–23,30] Mutations in *EGFR*, in particular, in-frame deletions of exon 19 and L858R and L861Q of exon 21, are strongly associated with never-smoking status, female gender and East Asian ethnicity, as well as predict favourable response to *EGFR* tyrosine kinase inhibitor.[2,12,13,17,22,23] On the other hand, mutations in *KRAS*, the most frequently mutated oncogene in lung adenocarcinoma, based on recent findings of the tumour-sequencing project, are strongly associated with development of adenocarcinomas linked to tobacco consumption.[2,16,17,20,21,23]

It has been suggested that the vast majority of AAH precursor lesions and adenocarcinomas *in situ* are associated with the TRU adenocarcinoma subtype that were found to express high levels of *TITF-1* and SP, leading to the conclusion that such adenocarcinomas are of the same lineage as terminal airway epithelial cells. In addition, it has been postulated that *EGFR* mutations are predominant in or specific to peripheral lung adenocarcinomas of the TRU subtype, which were suggested to arise from AAH lesions,[44,74,83] as 90 of 97 *EGFR* mutant adenocarcinomas were positive for *TITF-1*, and 91 of the 97 tumours were of the TRU subtype.[83] In addition, the hypothesis put forward that *EGFR* mutations are associated with or specific to the TRU subtype of lung adenocarcinomas is also in part due to the observation that the frequency of *EGFR* and *KRAS* mutations among AAH lesions, adenocarcinomas *in situ* and invasive adenocarcinomas is significantly different.[7,44] It was determined that whereas *KRAS* mutations decreased along adenocarcinoma progression, from 33% in AAH to 8% in adenocarcinomas, *EGFR* mutations were evenly distributed suggesting that *KRAS*-mutated AAH lesions rarely progress to adenocarcinomas. It is also important to mention, and as reviewed by Yatabe, that several studies performed gene expression profiling of lung adenocarcinomas and other histological subtypes of lung cancer and found that lung adenocarcinomas were heterogeneous and divided into different clusters.[44] Clusters with expression of *CC10* and features of alveolar signature such as *TITF-1* exhibited significantly better survival compared with adenocarcinomas in other clusters and comprised a higher frequency of *EGFR* mutations.

Mutations in the tyrosine-kinase domain of *EGFR* mutations were shown to be involved in the early pathogenesis of lung cancer, being identified in histologically normal epithelium of small bronchi and bronchioles adjacent to *EGFR* mutant adenocarcinomas[77] (discussed further in the next section of the review). *EGFR* mutations were detected in normal-appearing peripheral respiratory epithelium in 43% adenocarcinoma patients,[77] but not in patients without mutation in the tumour.[77] These findings may signify different cell types comprising the examined epithelia, which could represent sites of the cells of origin for *EGFR* mutant adenocarcinomas of the lung. Although the cell type having those mutations is unknown, our group has hypothesized that stem or progenitor cells of the bronchial and bronchiolar epithelium bear such mutations. It is also noteworthy that *EGFR* mutations were identified in only 3 of 40 AAH lesions examined[83,84] and were shown to be absent[22] or relatively infrequent in what was previously known as BAC of the lung.[84] These earlier observations support the argument that abnormalities of *EGFR* are not only relevant to the pathogenesis of alveolar-type lung

neoplasia but also may play drive peripheral lung adenocarcinoma from bronchiolar epithelium cells that are distinct from terminal respiratory and alveolar cells.[5,44] The different findings of *EGFR* mutation rates in AAH lesions may as well reflect the ethnicity (Asian vs Western) of the patients from which the lesions were isolated, as well as the standard practice of detection of small lesions such as AAH.

Field cancerization

Although the majority of lung cancer patients are current or former smokers (approximately 85%), a relatively small fraction of smokers (approximately 15%) develop primary lung tumours. Patients with early stage NSCLC, relative to other early stage malignancies, frequently exhibit recurrence or second primary tumour development after definitive treatment by surgery and removal of the original lung primary tumour. There is a large body of evidence that heavy smokers and patients who have survived an upper aerodigestive cancer comprise a high-risk population that may be targeted for early detection and chemoprevention efforts.[6] Although the risk of developing lung cancer decreases after smoking cessation, the risk never returns to baseline. Preneoplastic changes, namely dysplastic histological abnormalities, have been utilized as surrogate endpoints for chemopreventive studies. However, it was suggested that this 'shooting-in-the-dark' approach may explain the reasons behind the general failures of clinical chemoprevention studies.[3] It is also important to note that we are unable to predict which lifetime never-smokers or definitively treated never-smoker early-stage lung cancer patients will develop lung tumours or relapse. Therefore, novel approaches to identify the best population to be targeted for early detection and chemoprevention should be devised, and risk factors for lung cancer development or relapse need to be better defined. For these important purposes, a better understanding of the biology and molecular origins of lung cancer, for example, lung adenocarcinoma, is warranted. In this section of the review, we will describe the field cancerization phenomenon that herein refers to that occurring due to direct and indirect effects of smoking (field of injury) or independent of smoking in patients with and without cancer, with emphasis on aberrant molecular markers in histological normal epithelia that can be used to increase our understanding of lung cancer pathogenesis.

Smoking damaged epithelium and the lung field cancerization phenomenon

Earlier work by Danely Slaughter *et al*. in patients with oral cancer and oral premalignant lesions has suggested that histologically normal-appearing tissue adjacent to neoplastic and preneoplastic lesions display molecular abnormalities, some of which are in common with those in the tumours.[85] In 1961, a seminal report by Auerbach *et al*. suggested that cigarette smoke induces extensive histological changes in the bronchial epithelia in the lungs of smokers and that premalignant lesions are widespread and multifocal throughout the respiratory epithelium, suggestive of a field effect.[86] This phenomenon, coined 'field of cancerization', was later shown to be evident in various epithelial cell malignancies including lung cancer. Some degree of inflammation and inflammatory-related damage is almost invariably present in the central and peripheral airways of smokers and may precede the development of lung cancer.[87] Thus, the field of cancerization may be explained by both direct effect of tobacco carcinogens and initiation of inflammatory response. In this context, different theories for the origin of the field of cancerization or smoking-related field of injury have been put forward and will not be discussed here, as they have been nicely and extensively reviewed elsewhere by Steiling *et al*.[88]

Several studies focusing on the respiratory epithelium of lung cancer patients and smokers have demonstrated that multiple altered foci of bronchial epithelium are present throughout the airway.[27,28,89] A detailed analysis of histologically normal epithelium, and premalignant and malignant epithelia from lung SCC patients indicated that multiple, sequentially occurring allele-specific chromosomal deletions of LOH begin in dispersed clonally independent foci very early in the multi-stage pathogenesis of this smoking-related lung malignancy.[27,28] Notably, 31% of histologically normal epithelium and 42% of mildly abnormal (hyperplasia/metaplasia) specimens had clones of cells with allelic loss at one or more regions examined. Moreover, these molecular aberrations were also found in carcinomas *in situ* and SCC, and at a more advanced level.[27] Molecular changes involving LOH of chromosomal regions 3p (*DDUT* and *FHIT* genes), 9p (*CDKN2A*), genomic instability (increased microsatellite repeats) and *p16* methylation have all shown to commence in histologically normal or slightly abnormal tissue in SCC patients and in the sequence of pathogenesis of the disease.[5] As mentioned before, *KRAS* is the most mutated oncogene in lung adenocarcinomas.[37] Almost 15 years ago, Nelson *et al*. demonstrated that *KRAS* mutations are found in histologically normal lung tissue adjacent to lung tumours.[90] As will be discussed later, mutations in *EGFR* were also found in adjacent to tumour histologically normal epithelium.[77,78] Similar epigenetic and gene methylation patterns between tumours and adjacent histologically normal epithelia were described. An important study by Belinsky *et al*. reported aberrant promoter methylation of *p16*, which was described to be commonly methylated in lung tumours,[91] in at least one bronchial epithelial site from 44% of lung cancer patients examined.[92] Moreover, *p16* and death-associated protein kinase (DAPK) promoter methylation was observed frequently in bronchial epithelium from current and former smoker but not from never-smoker lung cancer patients and persisted after smoking cessation. Notably, 94% of lung tumours exhibited a concordant pattern of *p16* methylation with that in at least one bronchial epithelial site.[92]

The aforementioned molecular abnormalities were detected in histologically normal epithelia adjacent to archival surgically resected tumours from primary lung cancer patients. LOH and microsatellite alterations in multiple foci were also detected in distal histological normal bronchial epithelia of smokers without cancer.[93,94] Moreover, and importantly, these molecular abnormalities were detected in bronchial epithelia of cancer-free former smokers that appeared to have persisted for many years after smoking cessation. In addition, LOH was detected in DNA obtained from bronchial brushings of normal and abnormal lungs from patients undergoing diagnostic bronchoscopy and was detected in cells from the ipsilateral and contralateral lung.[95] Mutations in *TP53* were also described to occur in bronchial epithelia of cancer-free smokers in a widely dispersed manner.[96] Similar evidence also exists for promoter methylation and epigenetic changes in smoking-damaged lung epithelium of cancer-free patients. Methylation of various genes, including retinoic acid receptor 2 beta (*RAR-b2*), *H-cadherin*, *APC*, *p16* and *RASSFF1A* was described in bronchial epithelial cells of heavy smokers.[97] Moreover, methylation of *p16*, *GSTP1* and *DAPK* was reported to be evident in bronchial brushings of one third of the cancer-free smokers examined.[98] In the same study by Belinsky *et al*., as mentioned before, methylation of *p16* was detected in epithelia of cancer-free smokers.[92] A more detailed list of aberrant gene promoter methylation in lung cancer patients and cancer-free smokers is nicely summarized and explained in the review by Heller *et al*.[99]

Gene expression profiling of the lung field cancerization

High-throughput microarray profiling was used by several groups to study the transcriptome of lung airways. Hackett *et al*. utilized microarrays to study the expression of 44 anti-oxidant-related genes using bronchial brushings from cancer-free current smokers and never-smokers, and found significant upregulation of 16 of the antioxidant genes in the airways of smokers compared with non-smokers.[100] Later, Spira *et al*. described global alterations in gene expression between normal-appearing bronchial epithelium of healthy cancer-free smokers and that of non-smokers.[101] In addition, and in the reports by Spira *et al*. and Beane *et al*., irreversible changes in expression in airways of former smokers after years of smoking cessation were described that were thought to underlay the increased risk former smokers display, compared with never-smokers, for developing lung cancer long after they have discontinued smoking.[101,102] Schembri *et al*. also reported alterations in the expression of miRNA between large airways of current and never-smokers.[103] Notably, an 80-gene signature was derived from the transcriptome of large airway epithelial cells that can distinguish smokers without overt cancer from smokers with lung cancer and exhibited statistically significant utility characteristics of a lung cancer biomarker, despite originating from normal bronchial epithelia.[104] Moreover, the 80-gene signature, using publicly available microarray datasets, was able to distinguish lung tumours from corresponding normal lung tissues.[104] More recently, Gustafson *et al*. derived a phosphatidylinositol 3-kinase (*PI3K*) pathway activation signature using recombinant adenoviruses to express the 110α subunit of *PI3K* in primary human epithelial cells.[105] The same study then demonstrated that the *PI3K* pathway activation signature was elevated in cytologically normal bronchial airways of smokers with lung cancer and with dysplastic lesions.[105] Of substantial clinical importance, the study found that the signature was decreased in the airways of high-risk smokers whose dysplastic lesions regressed following treatment with the *PI3K* inhibitor myoinositol.[105]

Microarray and gene expression profiling methodologies were also used to demonstrate the wide anatomical spread of the lung field cancerization to epithelial regions that can be non-invasively sampled when devising approaches for early detection of lung cancer. Sridhar *et al*. highlighted common gene expression alterations in bronchial, nasal and buccal epithelia of smokers, in particular of various detoxification genes that perpetuate the field of cancerization due to tobacco consumption.[106] In addition, Zhang *et al*. identified 119 genes whose expression was affected by smoking similarly in both bronchial and nasal epithelium, including genes related to detoxification, oxidative stress and wound healing,[107] and the study by Boyle *et al*. highlighted significant similarities in expression changes between smokers and never-smokers in oral and bronchial epithelia.[108]

Lung adenocarcinoma field cancerization

To better understand the pathogenesis of *EGFR* mutant lung adenocarcinomas, Tang and colleagues investigated the presence of *EGFR* mutations in normal bronchial and bronchiolar epithelium adjacent to *EGFR* mutant tumours. As mentioned before, *EGFR* mutations were detected in histologically normal peripheral epithelia in 44% of lung adenocarcinoma patients with mutations but none in patients lacking mutations in the oncogene.[77] Moreover, the same study highlighted more frequent *EGFR* mutations in normal epithelium within the tumour (43%) than in adjacent sites (24%) suggests a localized field-effect phenomenon for this abnormality in the respiratory epithelium of the lung.[77] In addition, a higher frequency of mutations in cells obtained from small bronchi (35%) compared with bronchioles (18%) was detected.[77] More recently, *EGFR* protein overexpression, similar to mutation of the gene, also exhibited a localized field effect, as it was more frequent in normal bronchial epithelia sites within tumours than in sites adjacent to and distant from tumours.[78] Interestingly, *EGFR* copy-number alteration was not evident in normal bronchial epithelia, which is in accordance with findings that *EGFR* copy

number is relatively a late event in pathogenesis of adenocarcinomas.[7,78]

Field cancerization compartmentalization

The low frequency of molecular abnormalities detected in the centrally located bronchial respiratory epithelium in patients with peripheral lung adenocarcinomas, compared with specimens from patients with SCC and SCLC,[89] suggests the presence of two compartments in the lung with different degrees of smoking-related genetic damage. Thus, smokers who develop SCC have more smoking-related genetic damage in the respiratory epithelium of the central airway, whereas patients who develop adenocarcinoma have damage mainly in the peripheral airways (small bronchus, bronchioles and alveoli). While some molecular changes (e.g. inflammation and signalling pathways activation) have been detected throughout the lung airway and include both compartments (central and peripheral airway), other aberrations have been more frequently altered in either central (e.g. LOH, genetic instability evidenced by microsatellite repeats) or peripheral (e.g. *EGFR* mutations) airways.

Lineage-specific genes in lung cancer

The transformation of normal cells into tumourigenic counterparts is mediated by a complex array of intracellular signals, as well as genetic and epigenetic regulation. It has been suggested that lineage-specific genes, which play important roles in normal developmental processes such as organogenesis or tissue homeostasis and remain to be expressed or become amplified during an acquired pathological condition, are crucial for maintenance of the disease state.[32,109] Interestingly, lineage genes can discriminate different subtypes of the same cancer that rise from dissimilar cells/progenitors, for example, adenocarcinomas versus squamous tumours, and might offer new insights into crucial and therapeutically pliable tumour dependencies.[109] Various studies have highlighted the potential 'addiction' of tumour cells to aberrant and growth-promoting cell signalling mediated by lineage-specific oncogenes, for example, presence of the *BCR-ABL* fusion oncoprotein in chronic myelogenous leukemia,[110] mutations in the *KIT* oncogene in gastrointestinal stromal tumours,[111] amplification of the microphthalmia-associated transcriptional factor (*MITF*) in melanoma[112] and, more recently, amplification of *PAX8* in ovarian cancer.[113] Two lineage-specific oncogenes have been characterized in NSCLC. Recently, *TITF-1* amplification and protein expression were shown to be prevalent in lung adenocarcinomas and elicit growth-promoting signals in this malignancy.[34–36] The master ESC transcriptional factor *SOX2* was shown to be a member of the 3q locus (3q26.3) that is specifically amplified in lung and oesophageal squamous carcinomas.[32] These findings demonstrate that *TITF-1* and *SOX2* function as lineage-specific oncogenes in lung adenocarcinomas and SCC, respectively, and that targeting pathways downstream of those two master regulators may leverage new therapeutic strategies independently for each NSCLC subtype.

TITF-1

TITF-1 is a homeodomain-containing transactivating factor predominantly expressed in the terminal lung bronchioles and lung periphery in the developing and adult mouse.[114,115] In addition, *TITF-1* is crucial for branching morphogenesis during normal lung development[114–116] and transactivates the expression of the SP, such as SP-A, -B and -C, which are in turn typically expressed in the Clara cells and are important for the differentiation of alveolar type II pneumocyte cells in the peripheral lung.[117]

Several studies have demonstrated increased copy number and amplification of the 14q13.3 locus that harbours the *TITF-1* gene as well as paired box transcriptional factor family member 9 (*PAX9*) and *NKX2.8*.[34,36] It is postulated that *TITF-1* functions as a lineage-specific oncogene in lung adenocarcinoma as knockdown of *TITF-1* expression, in cells with amplification of the gene, by RNA interference results in lung adenocarcinoma cell-growth inhibition and apoptosis demonstrating a lineage-specific dependency of lung adenocarcinomas on *TITF-1*.[34–36] Kendall *et al*. demonstrated that co-amplified *TITF-1*, *PAX9* and *NKX2.8* exhibit oncogenic cooperation and cell prosurvival and proliferative properties.[118] Overexpression of both *TITF-1* and *NKX2.8* simultaneously in *BEAS-2B* immortalized human bronchial epithelial cells elicited the highest increase in cell colony growth compared with single-gene transfected cells.[118] Moreover, pathway gene signatures that overlap downstream of both *TITF-1* and *NKX2.8* defined lung adenocarcinoma patients with most dismal prognosis compared with signatures downstream of either transcriptional factor alone.[119] However, recently in *KRAS*(LSL-G12D/+);p53(flox/flox) mice, *TITF-1* was shown to suppress tumourigenesis and limit metastatic potential *in vivo*.[120]

Our group and others have demonstrated that *TITF-1* copy-number gain or amplification is associated with poor prognosis in NSCLC.[121,122] In contrast to the expected pro-survival properties of a cell-lineage oncogene and the association of *TITF-1* copy-number gain and amplification with poor survival, *TITF-1* protein expression by immunohistochemistry was shown to be a marker of favourable prognosis in NSCLC[122–125] including early stage (stage I) lung adenocarcinoma.[126] It is worthwhile to mention that *TITF-1* protein expression and *TITF-1* gene copy number were found to be associated with mutations in the *KRAS* and *EGFR* oncogenes, respectively.[122] As mutations in *EGFR* and *KRAS* occur almost mutually exclusively in lung adenocarcinomas[2] and were suggested to function in different lineages of lung adenocarcinomas,[109] it is possible that *TITF-1* expression is aberrantly differently controlled within different subsets of adenocarcinomas. It is also important to note that *TITF-1* copy-number gain was also

demonstrated in lung SCC.[34,122,125] It is plausible that *TITF-1* copy gain may only be a surrogate marker in SCC of another molecular defect in a gene nearby or within the 14q13.3 amplicon, for example, *NKX2.8* or *PAX9*. The significance of the infrequent copy number increase of *TITF-1* in lung SCC remains elusive.

SOX2

SOX2 was suggested to play key developmental roles in the formation of the lung, trachea and oesophagus based on its expression pattern in these tissues and organs.[127] Interestingly, *SOX2* was shown to be important for the morphogenesis of the trachea and oesophagus, and the differentiation of the oesophageal epithelium.[128] Moreover, the timing of *SOX2* expression in the foregut is tightly regulated, as it is only expressed in the main airways and non-branching bronchioles in the developing and adult mouse lung.[127,129,130] Heterozygote and homozygote transgenic mice with mutant *SOX2* have substantial defects in lung branching and morphogenesis during development.[129,130] Moreover, *SOX2* plays key roles in the maintenance of developing and adult tracheal cells evidenced by shorter and injured trachea in mice with knockout of both alleles of the transcriptional factor.[130] The numerous functions *SOX2* elicits in the differentiation of the conducting airways among other roles are reviewed in more detail by Whitsett and colleagues.[131] It is important to note that *SOX2* forms a core transcriptional factor complex with *OCT4* or *OCT1* and *TirNaNog*/*NANOG* that binds to enhancer sequences of various genes to regulate the inner cell mass or embryoblast within the blastocyst cavity in embryos.[132] Moreover, Boyer and colleagues demonstrated that *SOX2* along with *OCT4* and *NANOG* form a core regulatory transcriptional circuitry, signified by a *SOX2*/*OCT4*/*NANOG* expression signature, consisting of autoregulatory and feed-forward loops for the pluripotency and self-renewal of ESC.[133]

As mentioned earlier in the review, various studies have demonstrated that amplification of chromosomal region 3q (3q26.3) is almost specific to lung SCC.[61,63,64,66,67] The studies by Bass *et al*. and Hussenet and colleagues revealed that *SOX2* is amplified in this chromosomal region in lung SCC and squamous oesophageal cancers and promotes survival of SCC with amplification of this gene.[32,134] Subsequently, increased *SOX2* mRNA levels in lung SCC relative to adenocarcinomas was further evidenced by effective separation of both NSCLC subtypes by the previously characterized *OCT4*/*SOX2*/*NANOG* ESC expression signature,[133] following analysis of publicly available NSCLC microarray datasets.[33] In addition, SOX2 immunohistochemical protein expression was completely absent in lung adenocarcinoma pathogenesis, highly expressed in SCC development and significantly elevated in lung SCC relative to adenocarcinomas.[33] Interestingly, Maeir *et al*. later demonstrated that *SOX2* amplification was found in squamous carcinomas originating from other tissues and organs, such as those of the cervix, skin and penis.[135] It is noteworthy that SOX2 immunohistochemical protein expression in lung SCC and adenocarcinomas was also observed by other groups but in association with clinicopathological features including patient outcome. Interestingly, Wilbertz *et al*. reported the association of *SOX2* expression with favourable prognosis in lung SCC.[136] On the other hand, Sholl and colleagues demonstrated that *SOX2* immunohistochemical expression was an indicator of poor prognosis in lung adenocarcinomas.[137] Despite the equivocal associations of *SOX2* with lung cancer prognosis, various studies have highlighted tumour-promoting roles for this lineage-specific oncogene in lung cancer.[32,138,139]

McCaughan and colleagues specifically analysed 3q copy-number alteration in bronchial dysplasia of varying grades and severity and demonstrated that *SOX2* amplification was present in high-grade bronchial dysplasias but not in low-grade lesions and, importantly, was associated with clinical progression of high-grade preinvasive squamous lesions.[140] It is important to mention that Yuan *et al*. had found relatively high *SOX2* immunohistochemical protein expression in normal bronchial epithelia and alveolar bronchiolarization structures.[33] Congruent with the study by Yuan *et al*., the results by McCaughan and colleagues demonstrated the implication of *SOX2* in the early pathogenesis of lung SCC.[33,140] Given the high *SOX2* protein expression in histologically normal bronchial epithelia, amplification of *SOX2* in high-grade dysplasia may exacerbate signalling downstream of this transcriptional factor in the course of SCC development. It is unknown whether *SOX2* may be amplified in normal bronchial epithelia, in particular, those adjacent to lung SCC with increased dosage of the gene. The findings outlined earlier demonstrate that *SOX2* is another cell-lineage oncogene with dissimilar functions between SCC and lung adenocarcinomas.

FUTURE DIRECTIONS

Lung adenocarcinoma genomics

Studies addressing genomic profiles, including copy-number alterations and mutational spectrums, have substantially increased our understanding of the molecular make-up and biology of lung adenocarcinomas demonstrating that, genomically, this subtype of NSCLC is different from SCC. However, the heterogeneity within lung adenocarcinomas is still poorly understood. For example, it is unknown whether, for example, genomic copy-number alterations found in never-smoker adenocarcinomas are unique to this subtype or whether they are also found in smoker tumours. A large-scale side-by-side genomic analysis of never-smoker and smoker lung adenocarcinomas would shed light on copy-number alterations unique to both subtypes of lung adenocarcinomas. Moreover, it is not clear whether certain copy-number alterations can be clinically exploited for targeted therapy of lung adenocarcinoma. An important step in this direction was the demonstration by Yuan *et al*. that

lung adenocarcinomas with mutant *EGFR* and amplification of specific genes within the 7p region predict poor response to *EGFR* targeting tyrosine-kinase inhibitors.[70] It is tempting to speculate that an orthogonal study, largely encompassing both copy-number alterations and mutational spectrum and detecting focal amplification of oncogenes and loss of tumour-suppressor genes, would, for example, highlight potential targets of therapy in *EGFR*, *KRAS* and *ALK* wild-type lung adenocarcinomas for which there is an unmet need for therapeutic strategies.

Next-generation sequencing

Next-generation sequencing (NGS) technology, through whole-genome, whole-exome and whole-transcriptome approaches, holds great promise for providing invaluable insights into lung adenocarcinoma biology, diagnosis, prevention and therapy.[141] NGS enables the sequencing of expressed genes, exons and complete genomes providing data on levels of expression with a substantially larger dynamic range compared with array technology, sequence alterations, single nucleotide variations, as well as structural genomic aberrations.[141] A handful of studies have successfully applied NGS approaches to sequence one or two human lung tumour samples or cell lines demonstrating the feasibility of systematic, genome-wide characterization of rearrangements and alterations in complex human cancer genomes.[141–144] NGS analysis of a significant number of lung adenocarcinomas and/or NSCLC with characterized mutational status of known oncogenes (e.g. *EGFR* and *KRAS*) undoubtedly represents an important next step in furthering our comprehension of lung cancer biology. However, the application of NGS technology in clinical decision-making and personalized medicine is yet challenging.

Field cancerization and lung adenocarcinoma pathogenesis

Applying the same advanced high-throughput methodologies currently used in studying established tumours for the genetic analysis of lung adenocarcinoma preneoplasia and intraepithelial lesions, as well as histologically normal adjacent regions, is expected to expand our understanding of the biology of this prevalent disease. An important step in this direction was a recent study by Beane *et al.* in which RNA of bronchial airway epithelial cell brushings from healthy never-smokers and smokers with and without lung cancer was analysed by RNA sequencing.[145] The study highlighted transcripts whose expression was either not interrogated by or was not found to be significantly altered when using microarrays demonstrating that NGS, like in established lung tumours, has the potential to provide new insights into the biology of the airway field cancerization associated with smoking and lung cancer.[145]

Earlier findings demonstrated that centrally located lung SCC and peripherally located lung adenocarcinomas elicit and perpetuate differential effects on the airway epithelia. We believe that these effects overlap with those of the response of the host to tobacco exposure (reviewed by Steiling *et al.*[88]) but may be unique in several aspects. Changes in expression in the lung field of injury have shown to be similar in the large and small airways, and it is unknown whether they are associated with the development of the particular subtype of NSCLC. Addressing this question may be highly pertinent because both NSCLC subtypes display different genomic features, as previously discussed, and, therefore, are clinically managed by significantly dissimilar treatment strategies, let alone differences among various subtypes of lung adenocarcinomas. Moreover, a compartmental approach in studying the field of cancerization (Fig. 1) will shed

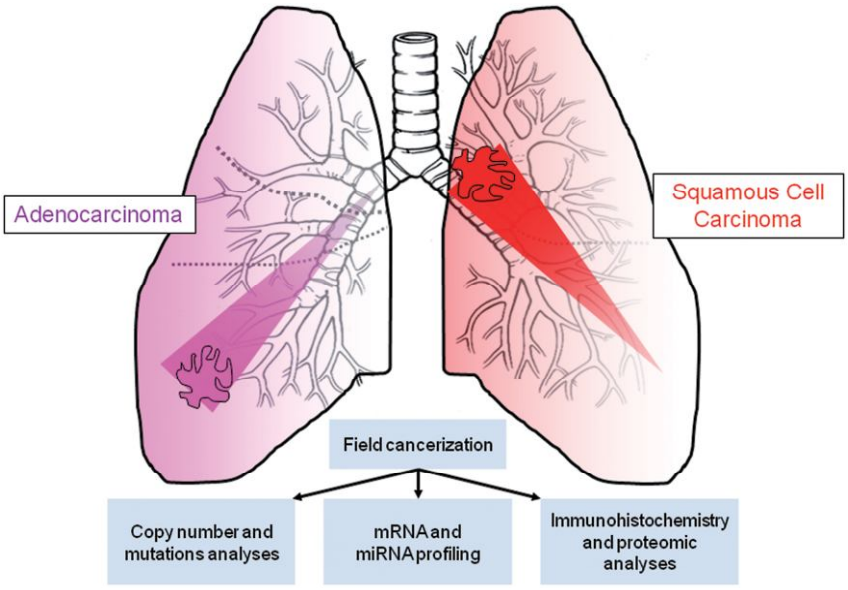


Figure 1 Molecular analysis of the lung field cancerization. It is unknown whether changes in expression in the lung field cancerization are associated with the development of a particular subtype of non-small cell lung cancer (NSCLC), that is, adenocarcinomas compared with squamous-cell carcinomas (SCC). Analysing local and distant field cancerization independently for lung adenocarcinomas and SCC may shed light on events common or unique to the molecular pathogenesis of the two major subtypes of NSCLC. Such a 'compartmental' approach in studying the field cancerization may unravel biomarkers that can guide personalized prevention strategies suitable for each different NSCLC subtype.

light on events in the early pathogenesis of lung adenocarcinomas versus SCC and unravel biomarkers that can be lineage specific and can guide personalized chemoprevention strategies suitable for each different NSCLC subtype, which may reduce the relatively high frequency of relapse of early stage patients.

PERSPECTIVE

Despite numerous efforts that have focused on increasing our understanding of the biology of lung adenocarcinomas, this subtype of NSCLC that is increasing in incidence compared with SCC, constitutes for approximately half of lung cancer deaths each year, which in turn comprise the biggest share of cancer-related deaths in the United States and worldwide. Compared with advances in targeted and personalized therapy of lung adenocarcinomas, little progress has been made in the tailored prevention of this fatal malignancy leading to a substantially decreased enthusiasm. This may change with the recent encouraging and significant findings of the NLST. Various molecular markers and expression classifiers previously described in the lung airways and in less invasive sites of the field cancerization, for example, nasal, sputum and exhaled breath condensates, can aid in selecting high-risk individuals best suited for CT screening for example. A comprehensive analysis of early molecular events in lung adenocarcinoma pathogenesis will undoubtedly unravel biomarkers that can, in the future, aid prevention through personalized strategies, deliver its longstanding promise to oppose this disease.

ACKNOWLEDGEMENTS

Funded in part by a Lung Cancer Research Foundation grant (HK) and DoD W81XWH-10-1-1007 (IIW).

REFERENCES

1 Jemal A, Bray F, Center MM *et al.* Global cancer statistics. *CA Cancer J. Clin.* 2010; **61**: 69–90.
2 Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N. Engl. J. Med.* 2008; **359**: 1367–80.
3 Gold KA, Kim ES, Lee JJ *et al.* The BATTLE to personalize lung cancer prevention through reverse migration. *Cancer Prev. Res. (Phila.)* 2011; **4**: 962–72.
4 Aberle DR, Adams AM, Berg CD *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* 2011; **365**: 395–409.
5 Wistuba II, Gazdar AF. Lung cancer preneoplasia. *Annu. Rev. Pathol.* 2006; **1**: 331–48.
6 Wistuba II. Genetics of preneoplasia: lessons from lung cancer. *Curr. Mol. Med.* 2007; **7**: 3–14.
7 Yatabe Y, Borczuk AC, Powell CA. Do all lung adenocarcinomas follow a stepwise progression? *Lung Cancer* 2011; **74**: 7–11.
8 Scagliotti G, Brodowicz T, Shepherd FA *et al.* Treatment-by-histology interaction analyses in three phase III trials show superiority of pemetrexed in nonsquamous non-small cell lung cancer. *J. Thorac. Oncol.* 2011; **6**: 64–70.
9 Scagliotti GV, Parikh P, von Pawel J *et al.* Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *J. Clin. Oncol.* 2008; **26**: 3543–51.
10 Kabat GC, Wynder EL. Lung cancer in nonsmokers. *Cancer* 1984; **53**: 1214–21.
11 Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung Cancer* 2001; **31**: 139–48.
12 Rudin CM, Avila-Tang E, Harris CC *et al.* Lung cancer in never smokers: molecular profiles and therapeutic implications. *Clin. Cancer Res.* 2009; **15**: 5646–61.
13 Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers—a different disease. *Nat. Rev. Cancer* 2007; **7**: 778–90.
14 Gazdar AF, Thun MJ. Lung cancer, smoke exposure, and sex. *J. Clin. Oncol.* 2007; **25**: 469–71.
15 Wakelee HA, Chang ET, Gomez SL *et al.* Lung cancer incidence in never smokers. *J. Clin. Oncol.* 2007; **25**: 472–8.
16 Le Calvez F, Mukeria A, Hunt JD *et al.* TP53 and KRAS mutation load and types in lung cancers in relation to tobacco smoke: distinct patterns in never, former, and current smokers. *Cancer Res.* 2005; **65**: 5076–83.
17 Mounawar M, Mukeria A, Le Calvez F *et al.* Patterns of EGFR, HER2, TP53, and KRAS mutations of p14arf expression in non-small cell lung cancers in relation to smoking history. *Cancer Res.* 2007; **67**: 5667–72.
18 Pao W, Miller V, Zakowski M *et al.* EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl Acad. Sci. U.S.A.* 2004; **101**: 13306–11.
19 Reynolds SH, Anna CK, Brown KC *et al.* Activated protooncogenes in human lung tumors from smokers. *Proc. Natl Acad. Sci. U.S.A.* 1991; **88**: 1085–9.
20 Ridanpaa M, Karjalainen A, Anttila S *et al.* Genetic alterations in p53 and K-ras in lung-cancer in relation to histopathology of the tumor and smoking history of the patient. *Int. J. Oncol.* 1994; **5**: 1109–17.
21 Rodenhuis S, Slebos RJ, Boot AJ *et al.* Incidence and possible clinical significance of K-ras oncogene activation in adenocarcinoma of the human lung. *Cancer Res.* 1988; **48**: 5738–41.
22 Shigematsu H, Lin L, Takahashi T *et al.* Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J. Natl Cancer Inst.* 2005; **97**: 339–46.
23 Tam IY, Chung LP, Suen WS *et al.* Distinct epidermal growth factor receptor and KRAS mutation patterns in non-small cell lung cancer patients with different tobacco exposure and clinicopathologic features. *Clin. Cancer Res.* 2006; **12**: 1647–53.
24 Travis WD, Garg K, Franklin WA *et al.* Bronchioloalveolar carcinoma and lung adenocarcinoma: the clinical importance and research relevance of the 2004 World Health Organization pathologic criteria. *J. Thorac. Oncol.* 2006; **1**: S13–9.
25 Motoi N, Szoke J, Riely GJ *et al.* Lung adenocarcinoma: modification of the 2004 WHO mixed subtype to include the major histologic subtype suggests correlations between papillary and micropapillary adenocarcinoma subtypes, EGFR mutations and gene expression analysis. *Am. J. Surg. Pathol.* 2008; **32**: 810–27.
26 Travis WD, Brambilla E, Noguchi M *et al.* International association for the study of lung cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *J. Thorac. Oncol.* 2011; **6**: 244–85.
27 Wistuba II, Behrens C, Milchgrub S *et al.* Sequential molecular abnormalities are involved in the multistage development of squamous cell lung carcinoma. *Oncogene* 1999; **18**: 643–50.
28 Wistuba II, Behrens C, Virmani AK *et al.* High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. *Cancer Res.* 2000; **60**: 1949–60.

29 Wistuba II, Behrens C, Virmani AK *et al.* Allelic losses at chromosome 8p21-23 are early and frequent events in the pathogenesis of lung cancer. *Cancer Res.* 1999; **59**: 1973–9.
30 Shigematsu H, Takahashi T, Nomura M *et al.* Somatic mutations of the HER2 kinase domain in lung adenocarcinomas. *Cancer Res.* 2005; **65**: 1642–6.
31 Stephens P, Hunter C, Bignell G *et al.* Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* 2004; **431**: 525–6.
32 Bass AJ, Watanabe H, Mermel CH *et al.* SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat. Genet.* 2009; **41**: 1238–42.
33 Yuan P, Kadara H, Behrens C *et al.* Sex determining region Y-Box 2 (SOX2) is a potential cell-lineage gene highly expressed in the pathogenesis of squamous cell carcinomas of the lung. *PLoS ONE* 2010; **5**: e9112.
34 Kwei KA, Kim YH, Girard L *et al.* Genomic profiling identifies TITF1 as a lineage-specific oncogene amplified in lung cancer. *Oncogene* 2008; **27**: 3635–40.
35 Tanaka H, Yanagisawa K, Shinjo K *et al.* Lineage-specific dependency of lung adenocarcinomas on the lung development regulator TTF-1. *Cancer Res.* 2007; **67**: 6007–11.
36 Weir BA, Woo MS, Getz G *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007; **450**: 893–8.
37 Ding L, Getz G, Wheeler DA *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008; **455**: 1069–75.
38 Barbacid M. Ras genes. *Annu. Rev. Biochem.* 1987; **56**: 779–827.
39 Campbell SL, Khosravi-Far R, Rossman KL *et al.* Increasing complexity of Ras signaling. *Oncogene* 1998; **17**: 1395–413.
40 Harris TJ, McCormick F. The molecular pathology of cancer. *Nat. Rev. Clin. Oncol.* 2010; **7**: 251–65.
41 Brunner TB, Hahn SM, Gupta AK *et al.* Farnesyltransferase inhibitors: an overview of the results of preclinical and clinical investigations. *Cancer Res.* 2003; **63**: 5656–68.
42 Hatzivassiliou G, Song K, Yen I *et al.* RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. *Nature* 2010; **464**: 431–5.
43 Haura EB, Ricart AD, Larson TG *et al.* A phase II study of PD-0325901, an oral MEK inhibitor, in previously treated patients with advanced non-small cell lung cancer. *Clin. Cancer Res.* 2010; **16**: 2450–7.
44 Yatabe Y. EGFR mutations and the terminal respiratory unit. *Cancer Metastasis Rev.* 2010; **29**: 23–36.
45 Lynch TJ, Bell DW, Sordella R *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 2004; **350**: 2129–39.
46 Paez JG, Janne PA, Lee JC *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004; **304**: 1497–500.
47 Soda M, Choi YL, Enomoto M *et al.* Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 2007; **448**: 561–6.
48 Kwak EL, Bang YJ, Camidge DR *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* 2010; **363**: 1693–703.
49 Koivunen JP, Mermel C, Zejnullahu K *et al.* EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer. *Clin. Cancer Res.* 2008; **14**: 4275–83.
50 Pao W, Girard N. New driver mutations in non-small-cell lung cancer. *Lancet Oncol.* 2011; **12**: 175–80.
51 Greulich H. The genomics of lung adenocarcinoma: opportunities for targeted therapies. *Genes Cancer* 2010; **1**: 1200–10.
52 Nigro JM, Baker SJ, Preisinger AC *et al.* Mutations in the p53 gene occur in diverse human tumour types. *Nature* 1989; **342**: 705–8.
53 Takahashi T, Nau MM, Chiba I *et al.* p53: a frequent target for genetic abnormalities in lung cancer. *Science* 1989; **246**: 491–4.
54 Hayashi N, Sugimoto Y, Tsuchiya E *et al.* Somatic mutations of the MTS (multiple tumor suppressor) 1/CDK4l (cyclin-dependent kinase-4 inhibitor) gene in human primary non-small cell lung carcinomas. *Biochem. Biophys. Res. Commun.* 1994; **202**: 1426–30.
55 Packenham JP, Taylor JA, White CM *et al.* Homozygous deletions at chromosome 9p21 and mutation analysis of p16 and p15 in microdissected primary non-small cell lung cancers. *Clin. Cancer Res.* 1995; **1**: 687–90.
56 Merlo A, Herman JG, Mao L *et al.* 5′CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat. Med.* 1995; **1**: 686–92.
57 Sun Y, Ren Y, Fang Z *et al.* Lung adenocarcinoma from East Asian never-smokers is a disease largely defined by targetable oncogenic mutant kinases. *J. Clin. Oncol.* 2010; **28**: 4616–20.
58 Pinkel D, Segraves R, Sudar D *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* 1998; **20**: 207–11.
59 Pollack JR, Perou CM, Alizadeh AA *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 1999; **23**: 41–6.
60 Kim TM, Yim SH, Lee JS *et al.* Genome-wide screening of genomic alterations and their clinicopathologic implications in non-small cell lung cancers. *Clin. Cancer Res.* 2005; **11**: 8235–42.
61 Tonon G, Wong KK, Maulik G *et al.* High-resolution genomic profiles of human lung cancer. *Proc. Natl Acad. Sci. U.S.A.* 2005; **102**: 9625–30.
62 Balsara BR, Testa JR. Chromosomal imbalances in human lung cancer. *Oncogene* 2002; **21**: 6877–83.
63 Bjorkqvist AM, Husgafvel-Pursiainen K, Anttila S *et al.* DNA gains in 3q occur frequently in squamous cell carcinoma of the lung, but not in adenocarcinoma. *Genes Chromosomes Cancer* 1998; **22**: 79–82.
64 Luk C, Tsao MS, Bayani J *et al.* Molecular cytogenetic analysis of non-small cell lung carcinoma by spectral karyotyping and comparative genomic hybridization. *Cancer Genet. Cytogenet.* 2001; **125**: 87–99.
65 Pei J, Balsara BR, Li W *et al.* Genomic imbalances in human lung adenocarcinomas and squamous cell carcinomas. *Genes Chromosomes Cancer* 2001; **31**: 282–7.
66 Petersen I, Bujard M, Petersen S *et al.* Patterns of chromosomal imbalances in adenocarcinoma and squamous cell carcinoma of the lung. *Cancer Res.* 1997; **57**: 2331–5.
67 Massion PP, Kuo WL, Stokoe D *et al.* Genomic copy number analysis of non-small cell lung cancer using array comparative genomic hybridization: implications of the phosphatidylinositol 3-kinase pathway. *Cancer Res.* 2002; **62**: 3636–40.
68 Job B, Bernheim A, Beau-Faller M *et al.* Genomic aberrations in lung adenocarcinoma in never smokers. *PLoS ONE* 2010; **5**: e15145.
69 Soh J, Okumura N, Lockwood WW *et al.* Oncogene mutations, copy number gains and mutant allele specific imbalance (MASI) frequently occur together in tumor cells. *PLoS ONE* 2009; **4**: e7464.
70 Yuan S, Yu SL, Chen HY *et al.* Clustered genomic alterations in chromosome 7p dictate outcomes and targeted treatment responses of lung adenocarcinoma with EGFR-activating mutations. *J. Clin. Oncol.* 2011; **29**: 3435–42.
71 Bhattacharjee A, Richards WG, Staunton J *et al.* Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl Acad. Sci. U.S.A.* 2001; **98**: 13790–5.
72 Garber ME, Troyanskaya OG, Schluens K *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl Acad. Sci. U.S.A.* 2001; **98**: 13784–9.
73 Hayes DN, Monti S, Parmigiani G *et al.* Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J. Clin. Oncol.* 2006; **24**: 5079–90.
74 Yatabe Y, Mitsudomi T, Takahashi T. TTF-1 expression in pulmonary adenocarcinomas. *Am. J. Surg. Pathol.* 2002; **26**: 767–73.

75 Westra WH. Early glandular neoplasia of the lung. *Respir. Res.* 2000; **1**: 163–9.
76 Chapman AD, Kerr KM. The association between atypical adenomatous hyperplasia and primary lung cancer. *Br. J. Cancer* 2000; **83**: 632–6.
77 Tang X, Shigematsu H, Bekele BN *et al.* EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. *Cancer Res.* 2005; **65**: 7568–72.
78 Tang X, Varella-Garcia M, Xavier AC *et al.* Epidermal growth factor receptor abnormalities in the pathogenesis and progression of lung adenocarcinomas. *Cancer Prev. Res. (Phila.)* 2008; **1**: 192–200.
79 Kitamura H, Kameda Y, Ito T *et al.* Atypical adenomatous hyperplasia of the lung. Implications for the pathogenesis of peripheral lung adenocarcinoma. *Am. J. Clin. Pathol.* 1999; **111**: 610–22.
80 Westra WH, Baas IO, Hruban RH *et al.* K-ras oncogene activation in atypical alveolar hyperplasias of the human lung. *Cancer Res.* 1996; **56**: 2224–8.
81 Chiosea S, Jelezcova E, Chandran U *et al.* Overexpression of Dicer in precursor lesions of lung adenocarcinoma. *Cancer Res.* 2007; **67**: 2345–50.
82 Selamat SA, Galler JS, Joshi AD *et al.* DNA methylation changes in atypical adenomatous hyperplasia, adenocarcinoma in situ, and lung adenocarcinoma. *PLoS ONE* 2011; **6**: e21443.
83 Yatabe Y, Kosaka T, Takahashi T *et al.* EGFR mutation is specific for terminal respiratory unit type adenocarcinoma. *Am. J. Surg. Pathol.* 2005; **29**: 633–9.
84 Yoshida Y, Shibata T, Kokubu A *et al.* Mutations of the epidermal growth factor receptor gene in atypical adenomatous hyperplasia and bronchioloalveolar carcinoma of the lung. *Lung Cancer* 2005; **50**: 1–8.
85 Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* 1953; **6**: 963–8.
86 Auerbach O, Stout AP, Hammond EC *et al.* Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. *N. Engl. J. Med.* 1961; **265**: 253–67.
87 Sin DD, Man SF, McWilliams A *et al.* Progression of airway dysplasia and C-reactive protein in smokers at high risk of lung cancer. *Am. J. Respir. Crit. Care Med.* 2006; **173**: 535–9.
88 Steiling K, Ryan J, Brody JS *et al.* The field of tissue injury in the lung and airway. *Cancer Prev. Res. (Phila.)* 2008; **1**: 396–403.
89 Wistuba II, Berry J, Behrens C *et al.* Molecular changes in the bronchial epithelium of patients with small cell lung cancer. *Clin. Cancer Res.* 2000; **6**: 2604–10.
90 Nelson MA, Wymer J, Clements N, Jr. Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers. *Cancer Lett.* 1996; **103**: 115–21.
91 Belinsky SA, Nikula KJ, Palmisano WA *et al.* Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc. Natl Acad. Sci. U.S.A.* 1998; **95**: 11891–6.
92 Belinsky SA, Palmisano WA, Gilliland FD *et al.* Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res.* 2002; **62**: 2370–7.
93 Mao L, Lee JS, Kurie JM *et al.* Clonal genetic alterations in the lungs of current and former smokers. *J. Natl Cancer Inst.* 1997; **89**: 857–62.
94 Wistuba II, Lam S, Behrens C *et al.* Molecular damage in the bronchial epithelium of current and former smokers. *J. Natl Cancer Inst.* 1997; **89**: 1366–73.
95 Powell CA, Klares S, O'Connor G *et al.* Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer. *Clin. Cancer Res.* 1999; **5**: 2025–34.
96 Franklin WA, Gazdar AF, Haney J *et al.* Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. *J. Clin. Invest.* 1997; **100**: 2133–7.
97 Zochbauer-Muller S, Lam S, Toyooka S *et al.* Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. *Int. J. Cancer* 2003; **107**: 612–6.
98 Soria JC, Rodriguez M, Liu DD *et al.* Aberrant promoter methylation of multiple genes in bronchial brush samples from former cigarette smokers. *Cancer Res.* 2002; **62**: 351–5.
99 Heller G, Zielinski CC, Zochbauer-Muller S. Lung cancer: from single-gene methylation to methylome profiling. *Cancer Metastasis Rev.* 2010; **29**: 95–107.
100 Hackett NR, Heguy A, Harvey BG *et al.* Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers. *Am. J. Respir. Cell Mol. Biol.* 2003; **29**: 331–43.
101 Spira A, Beane J, Shah V *et al.* Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc. Natl Acad. Sci. U.S.A.* 2004; **101**: 10143–8.
102 Beane J, Sebastiani P, Liu G *et al.* Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol.* 2007; **8**: R201.
103 Schembri F, Sridhar S, Perdomo C *et al.* MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proc. Natl Acad. Sci. U.S.A.* 2009; **106**: 2319–24.
104 Spira A, Beane JE, Shah V *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat. Med.* 2007; **13**: 361–6.
105 Gustafson AM, Soldi R, Anderlind C *et al.* Airway PI3K pathway activation is an early and reversible event in lung cancer development. *Sci. Transl. Med.* 2010; **2**: 26ra5.
106 Sridhar S, Schembri F, Zeskind J *et al.* Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* 2008; **9**: 259.
107 Zhang X, Sebastiani P, Liu G *et al.* Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. *Physiol. Genomics* 2010; **41**: 1–8.
108 Boyle JO, Gumus ZH, Kacker A *et al.* Effects of cigarette smoke on the human oral mucosal transcriptome. *Cancer Prev. Res. (Phila.)* 2010; **3**: 266–78.
109 Garraway LA, Sellers WR. Lineage dependency and lineage-survival oncogenes in human cancer. *Nat. Rev. Cancer* 2006; **6**: 593–602.
110 Kantarjian H, Sawyers C, Hochhaus A *et al.* Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. *N. Engl. J. Med.* 2002; **346**: 645–52.
111 Demetri GD, von Mehren M, Blanke CD *et al.* Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. *N. Engl. J. Med.* 2002; **347**: 472–80.
112 Garraway LA, Widlund HR, Rubin MA *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* 2005; **436**: 117–22.
113 Cheung HW, Cowley GS, Weir BA *et al.* Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc. Natl Acad. Sci. U.S.A.* 2011; **108**: 12372–7.
114 Minoo P, Su G, Drum H *et al.* Defects in tracheoesophageal and lung morphogenesis in Nkx2.1(-/-) mouse embryos. *Dev. Biol.* 1999; **209**: 60–71.
115 Yuan B, Li C, Kimura S *et al.* Inhibition of distal lung morphogenesis in Nkx2.1(-/-) embryos. *Dev. Dyn.* 2000; **217**: 180–90.
116 Kimura S, Hara Y, Pineau T *et al.* The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev.* 1996; **10**: 60–9.
117 Ikeda K, Clark JC, Shaw-White JR *et al.* Gene structure and expression of human thyroid transcription factor-1 in respiratory epithelial cells. *J. Biol. Chem.* 1995; **270**: 8108–14.
118 Kendall J, Liu Q, Bakleh A *et al.* Oncogenic cooperation and coamplification of developmental transcription factor genes in lung cancer. *Proc. Natl Acad. Sci. U.S.A.* 2007; **104**: 16663–8.

119 Hsu DS, Acharya CR, Balakumaran BS *et al.* Characterizing the developmental pathways TTF-1, NKX2-8, and PAX9 in lung cancer. *Proc. Natl Acad. Sci. U.S.A.* 2009; **106**: 5312–7.
120 Winslow MM, Dayton TL, Verhaak RG *et al.* Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* 2011; **473**: 101–4.
121 Barletta JA, Perner S, Iafrate AJ *et al.* Clinical significance of TTF-1 protein expression and TTF-1 gene amplification in lung adenocarcinoma. *J. Cell. Mol. Med.* 2009; **13**: 1977–86.
122 Tang X, Kadara H, Behrens C *et al.* Abnormalities of the TITF-1 lineage-specific oncogene in NSCLC: implications in lung cancer pathogenesis and prognosis. *Clin. Cancer Res.* 2011; **17**: 2434–43.
123 Barlesi F, Pinot D, Legoffic A *et al.* Positive thyroid transcription factor 1 staining strongly correlates with survival of patients with adenocarcinoma of the lung. *Br. J. Cancer* 2005; **93**: 450–2.
124 Berghmans T, Paesmans M, Mascaux C *et al.* Thyroid transcription factor 1—a new prognostic factor in lung cancer: a meta-analysis. *Ann. Oncol.* 2006; **17**: 1673–6.
125 Perner S, Wagner PL, Soltermann A *et al.* TTF1 expression in non-small cell lung carcinoma: association with TTF1 gene amplification and improved survival. *J. Pathol.* 2009; **217**: 65–72.
126 Anagnostou VK, Syrigos KN, Bepler G *et al.* Thyroid transcription factor 1 is an independent prognostic factor for patients with stage I lung adenocarcinoma. *J. Clin. Oncol.* 2009; **27**: 271–8.
127 Ishii Y, Rex M, Scotting PJ *et al.* Region-specific expression of chicken Sox2 in the developing gut and lung epithelium: regulation by epithelial-mesenchymal interactions. *Dev. Dyn.* 1998; **213**: 464–75.
128 Que J, Okubo T, Goldenring JR *et al.* Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. *Development* 2007; **134**: 2521–31.
129 Gontan C, de Munck A, Vermeij M *et al.* Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation. *Dev. Biol.* 2008; **317**: 296–309.
130 Que J, Luo X, Schwartz RJ *et al.* Multiple roles for Sox2 in the developing and adult mouse trachea. *Development* 2009; **136**: 1899–907.
131 Whitsett JA, Haitchi HM, Maeda Y. Intersections between pulmonary development and disease. *Am. J. Respir. Crit. Care Med.* 2011; **184**: 401–6.
132 Boiani M, Scholer HR. Regulatory networks in embryo-derived pluripotent stem cells. *Nat. Rev. Mol. Cell Biol.* 2005; **6**: 872–84.
133 Boyer LA, Lee TI, Cole MF *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 2005; **122**: 947–56.
134 Hussenet T, Dali S, Exinger J *et al.* SOX2 is an oncogene activated by recurrent 3q26.3 amplifications in human lung squamous cell carcinomas. *PLoS ONE* 2010; **5**: e8960.
135 Maier S, Wilbertz T, Braun M *et al.* SOX2 amplification is a common event in squamous cell carcinomas of different organ sites. *Hum. Pathol.* 2011; **42**: 1078–88.
136 Wilbertz T, Wagner P, Petersen K *et al.* SOX2 gene amplification and protein overexpression are associated with better outcome in squamous cell lung cancer. *Mod. Pathol.* 2011; **24**: 944–53.
137 Sholl LM, Barletta JA, Yeap BY *et al.* Sox2 protein expression is an independent poor prognostic indicator in stage I lung adenocarcinoma. *Am. J. Surg. Pathol.* 2010; **34**: 1193–8.
138 Lu Y, Futtner C, Rock JR *et al.* Evidence that SOX2 overexpression is oncogenic in the lung. *PLoS ONE* 2010; **5**: e11022.
139 Xiang R, Liao D, Cheng T *et al.* Downregulation of transcription factor SOX2 in cancer stem cells suppresses growth and metastasis of lung cancer. *Br. J. Cancer* 2011; **104**: 1410–7.
140 McCaughan F, Pole JC, Bankier AT *et al.* Progressive 3q amplification consistently targets SOX2 in preinvasive squamous lung cancer. *Am. J. Respir. Crit. Care Med.* 2010; **182**: 83–91.
141 Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.* 2010; **11**: 685–96.
142 Campbell PJ, Stephens PJ, Pleasance ED *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* 2008; **40**: 722–9.
143 Lee W, Jiang Z, Liu J *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 2010; **465**: 473–7.
144 Pleasance ED, Stephens PJ, O'Meara S *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 2010; **463**: 184–90.
145 Beane J, Vick J, Schembri F *et al.* Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. *Cancer Prev. Res. (Phila.)* 2011; **4**: 803–17.

Giles F. Filley Lecture

Field Cancerization in Non–Small Cell Lung Cancer
Implications in Disease Pathogenesis

Humam Kadara[1] and Ignacio I. Wistuba[1,2]

[1]Department of Thoracic/Head and Neck Medical Oncology, and [2]Department of Pathology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

Lung cancer, of which non–small cell lung cancer (NSCLC) composes the majority, is the leading cause of cancer-related deaths in the United States and worldwide. NSCLCs are tumors with complex biology that we have recently started to understand with the advent of various histological, transcriptomic, genomic, and proteomic technologies. However, the histological and molecular pathogenesis of this malignancy, in particular of adenocarcinomas, is still largely unknown. Earlier studies have highlighted a field cancerization phenomenon in which histologically normal-appearing tissue adjacent to neoplastic and pre-neoplastic lesions display molecular abnormalities, some of which are in common with those in the tumors. This review will summarize advances in understanding the field cancerization phenomenon and the potential relevance of this knowledge to gain important and novel insights into the molecular pathogenesis of NSCLC as well as to subsequent development of biomarkers for early detection of lung cancers and possibly personalized prevention.

Keywords: lung cancer; field cancerization; pathogenesis; airway epithelium

Lung cancer is the leading cause of cancer deaths in the United States and worldwide in both developing and developed regions (1). The high mortality of this disease is in part due to our lacking knowledge of the molecular mechanisms governing lung cancer pathogenesis as well as the late diagnosis of the majority of lung cancers after regional or distant spread of the malignancy (2). Non–small cell lung cancer (NSCLC) represents the majority of diagnosed lung cancers (2) and is mainly composed of squamous cell carcinomas (SCCs) and lung adenocarcinomas (2, 3). Several major differences exist between adenocarcinomas and SCCs. For example, compared with SCCs that arise from the major bronchi and are mainly centrally located, pulmonary adenocarcinomas arise from small bronchi, bronchioles, or alveolar epithelial cells and are typically peripherally located, as reviewed elsewhere (2–5). Moreover, whereas SCC pathogenesis is strongly linked to smoking, lung adenocarcinoma is the more common histological subtype in never-smoker patients (6–9). Although the sequence of lesions in the pathogenesis leading to SCCs is well described, little is known about the sequential development of adenocarcinomas. Moreover, we are still lacking in our knowledge of differential mechanisms of molecular pathogenesis among both subtypes of NSCLC.

(*Received in original form January 10, 2012; accepted in final form March 1, 2012*)

Supported by Department of Defense grant W81XWH-10-1-1007 (I.I.W.).

Correspondence and requests for reprints should be addressed to Ignacio I. Wistuba, M.D., Departments of Thoracic/Head and Neck Medical Oncology and Pathology, the University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Unit 085, Houston, TX 77030. E-mail: iiwistuba@mdanderson.org.

Proc Am Thorac Soc Vol 9, Iss. 2, pp 38–42, May 1, 2012

DOI: 10.1513/pats.201201-004MS
Internet address: www.atsjournals.org

In light of the postulated period of time and multiple stages required for the development of overt epithelial tumors, it is plausible to assume that early diagnosis of lung cancer or intraepithelial lesions coupled with effective prevention strategies will reduce the significant health burden associated with this disease (10). Despite recent encouraging findings from the National Lung Screening Trial (NLST) (11), early detection and prevention of lung cancer is challenging due to the lack of biomarkers for early diagnosis of the disease and to the presence of multiple neoplastic molecular pathways that mediate lung carcinogenesis. Earlier studies have highlighted a field cancerization phenomenon in which histologically normal-appearing tissue adjacent to neoplastic and pre-neoplastic lesions displays molecular abnormalities, some of which are in common with those in the tumors. It is plausible to assume that understanding early events in tumor development that commence in histologically normal epithelium would pave the way for unmet effective and personalized strategies for lung cancer prevention and treatment. This review mainly summarizes advances in understanding the field cancerization phenomenon and the potential relevance of this knowledge to gain important and novel insights into the molecular pathogenesis of NSCLC as well as to subsequent development of biomarkers for early detection of lung cancers and possibly personalized prevention.

NSCLC PRE-NEOPLASIA AND MOLECULAR PATHOGENESIS

From biological and histopathological perspectives, NSCLC is a complex malignancy that develops through multiple pre-neoplastic pathways. Lung adenocarcinoma, a major subtype of NSCLC, has been increasing in incidence globally in both smokers and nonsmokers (9), with a concurrent decrease in SCC frequency. It has been suggested that the increasing incidence of lung adenocarcinomas compared with SCCs is in part due to the change in the type of cigarettes used (lower nicotine and tar) and smoking habits and behavior (7). Anatomical differences in the location of diagnosed lung adenocarcinomas and SCCs strongly suggest that both NSCLC subtypes develop through different histopathological and molecular pathways and have different cells of origin; however, the specific respiratory epithelial cell type from which each lung cancer type develops has not been established with certainty (3). However, it is noteworthy that Clara cells and the type II pneumocytes are believed to be the progenitor cells of the peripheral airways, and peripherally arising adenocarcinomas often express markers of these cell types (12, 13).

The multistage stepwise fashion of tumor development has been demonstrated in various anatomical organs exemplified by the operational stages occurring during skin carcinogenesis (14). Carcinogenesis of the skin is initiated by a carcinogen-induced mutational event, promoted by clonal outgrowth, which may be dependent on tumor promoters, followed by progression of premalignant lesions (e.g., papillomas) and their conversion to malignant tumors (14).

Importantly, earlier genetic studies indicate that nonrandom, sequential chromosomal aberrations (trisomies of chromosomes 6 and 7) are associated with premalignant progression of mouse skin papillomas (14–17). The sequence of histopathological and molecular changes in bronchial epithelia that precede the development of lung SCCs have been characterized, demonstrating that sequentially occurring allele-specific molecular changes commence in dispersed foci signifying a multistage fashion of squamous lung cancer pathogenesis (4, 18). At least in this subtype of NSCLC, it has been shown that specific genomic alterations (mainly 3p21, 3p22–24, 3p25, and 9p21) occur in histologically normal bronchial epithelia from resected specimens (18). Moreover, and notably, these alterations persisted in hyperplasias, dysplasias, carcinomas *in situ*, and tumors that exhibited different commencing genomic aberrations (18). In addition, loss of heterozygosity (LOH) in the 3p region was demonstrated in normal bronchial epithelia of cancer-free smokers, further highlighting the early role this specific genomic alteration exerts in lung cancer pathogenesis (19, 20). It is noteworthy that McCaughan and colleagues demonstrated that no low-grade lesions, but all high-grade lesions, exhibited 3q amplification targeting the sex determining region Y-box 2 (*SOX2*) lineage-specific oncogene (21). These previous findings and reports, through highlighting associations between histopathological sequences and specific molecular aberrations, pinpoint to a multistage and multistep manner of lung carcinogenesis.

On the other hand, one cannot neglect the alternative hypothesis that the sequence of genetic and epigenetic alterations is irrelevant to lung cancer pathogenesis, but rather the accumulation of molecular abnormalities beyond a certain threshold mediates development of the malignant phenotype. It has been suggested that at least two molecular pathways, the Kirsten rat sarcoma viral oncogene (*KRAS*) and epidermal growth factor receptor (*EGFR*) pathways, are involved in the development of smoker and never-smoker adenocarcinomas, respectively (2, 9). Moreover, and as reviewed by Yatabe and colleagues, atypical adenomatous hyperplasias (AAHs), which are considered to be precursor lesions for peripheral lung adenocarcinomas (3, 5) and the only sequence of morphologic change identified so far for the development of invasive lung adenocarcinomas, exhibit *KRAS* mutations more frequently than invasive adenocarcinomas (5). Conversely, our group has previously demonstrated that *EGFR* mutations commence in histologically normal bronchial epithelia adjacent to lung adenocarcinomas and precede copy number increase of the oncogene (22, 23). *EGFR* mutations also are persistent throughout the different phases of lung adenocarcinoma development (5), which harbor different genomic alterations (24). It is plausible to surmise that only after increasing our knowledge of the pre-neoplastic changes as well as the corresponding molecular abnormalities leading to the development of lung adenocarcinomas would we then be able to more confidently determine whether adenocarcinomas follow a linear progression mechanism or not (5). However, and based on the aforementioned previously reported observations by our group and others, we believe that it is not counterintuitive to speculate that development of lung malignant phenotype, including that of adenocarcinomas, is due to stepwise, sequence-specific, and multistage molecular pathogenesis as well as accumulation and combination of genetic and epigenetic abnormalities.

FIELD CANCERIZATION

Smoking-Damaged Epithelium and the Field Cancerization Phenomenon

Although the majority of lung cancer patients are current or former smokers, a relatively small proportion of these smokers (approximately 15%) develop primary lung tumors. Patients with early-stage NSCLC commonly exhibit recurrence or second primary tumor development after definitive treatment by resective surgery. There is a large body of evidence that heavy smokers and patients who have survived lung cancer compose a high-risk population that may be targeted for early detection and chemoprevention efforts (4). Although the risk of developing lung cancer decreases after smoking cessation, the risk never returns to baseline. Pre-neoplastic changes have been used as surrogate endpoints for chemopreventive studies. However, it was suggested that this "shooting in the dark" approach may explain the reasons behind the general failures of clinical chemoprevention studies (10). Therefore, novel approaches to identify the best population to be targeted for early detection and chemoprevention should be devised, and risk factors for lung cancer development or relapse need to be better defined. For these important purposes, a better understanding of the biology and molecular origins of lung cancer is warranted.

Earlier work by Danely Slaughter in patients with oral cancer and oral premalignant lesions has suggested that histologically normal-appearing tissue adjacent to neoplastic and pre-neoplastic lesions display molecular abnormalities, some of which are in common with those in the tumors (25). In 1961, a seminal report by Auerbach and colleagues suggested that cigarette smoke induces extensive histological changes in the bronchial epithelia in the lungs of smokers and that premalignant lesions are widespread and multifocal throughout the respiratory epithelium, suggestive of a field effect (26). This phenomenon, coined "field of cancerization," was later shown to be evident in various epithelial cell malignancies, including lung cancer. Some degree of inflammation and inflammatory-related damage is almost invariably present in the central and peripheral airways of smokers and may precede the development of lung cancer (4). Thus, the field of cancerization may also be explained by both the direct effect of tobacco carcinogens and the initiation of inflammatory response. In this context, different theories for the origin of the field cancerization or smoking-related field of injury have been put forward and extensively reviewed elsewhere by Steiling and coworkers (27).

Multiple altered foci of bronchial epithelium are present throughout the airway in patients with lung cancer and smokers (18, 28, 29). As mentioned before, detailed analysis of histologically normal, premalignant, and malignant epithelia from patients with lung SCC indicated that multiple, sequentially occurring allele-specific chromosomal deletions of LOH commence in clonally independent foci early in the multistage pathogenesis of SCCs (18, 28). Notably, 31% percent of histologically normal epithelium and 42% of mildly abnormal (hyperplasia/metaplasia) specimens had clones of cells with allelic loss at one or more regions examined. Moreover, these molecular aberrations were also found in carcinomas *in situ* and SCCs and at a more advanced level (18). Molecular changes involving LOH of chromosomal regions 3p (e.g., fragile histidine triad gene/*FHIT*), 9p (e.g., *CDKN2A/p16*), genomic instability (increased microsatellite repeats), and *p16* methylation have been demonstrated in histologically normal bronchial epithelia in patients with SCC and in the sequence of pathogenesis of the disease (3). Moreover, Nelson and colleagues demonstrated that *KRAS* is also mutated in histologically normal lung tissue adjacent to lung tumors (30). In addition, similar epigenetic and gene methylation patterns between tumors and adjacent histologically normal epithelia were described. Belinsky and colleagues reported aberrant promoter methylation of *p16*, which was described to be commonly methylated in lung tumors (31), in at least one bronchial epithelial site from 44% of lung cancer cases examined (32). Moreover, *p16* and death-associated protein kinase (*DAPK*) promoter

methylation were frequently observed in bronchial epithelium from smoker but not from never-smoker patients with lung cancer and persisted after smoking cessation (32).

The aforementioned molecular abnormalities were detected in histologically normal epithelia adjacent to archival surgically resected tumors from patients with primary lung cancer. LOH and microsatellite alterations in multiple foci were also detected in distal histological normal bronchial epithelia of smokers without cancer (19, 20). Moreover, and importantly, these molecular abnormalities were detected in bronchial epithelia of cancer-free former smokers and appeared to have persisted for many years after smoking cessation. In addition, LOH was detected in DNA obtained from bronchial brushings of normal and abnormal lungs from patients undergoing diagnostic bronchoscopy and was detected in cells from the ipsilateral and contralateral lung (33). Mutations in *TP53* were also described to occur in bronchial epithelia of cancer-free smokers in a widely dispersed manner (34). Similar evidence also exists for promoter methylation and epigenetic changes in smoking-damaged lung epithelium of cancer-free patients. Methylation of various genes, including retinoic acid receptor 2 β (*RAR-β2*), *H-cadherin*, adenomatous polyposis coli (*APC*), *p16*, and Ras association (RalGDS/AF-6) domain family member 1 (*RASSFF1A*) has been described in bronchial epithelial cells of heavy smokers (35). Moreover, methylation of *p16*, glutathione S-transferase pi 1 (*GSTP1*), and *DAPK* was reported to be evident in bronchial brushings of one-third of cancer-free smokers examined (36). In the study by Belinsky and colleagues, as mentioned before, methylation of *p16* was detected in epithelia of cancer-free smokers (32). A more detailed list of aberrant gene promoter methylation in patients with lung cancer and cancer-free smokers is well summarized and explained in the review by Heller and coworkers (37).

Transcriptomic Studies of Lung Field Cancerization

High-throughput microarray profiling was shown to be useful to study the transcriptome of lung airways. Hackett and colleagues studied the expression of 44 antioxidant-related genes using bronchial brushings from cancer-free current smokers and never-smokers and found significant up-regulation of 16 of the antioxidant genes in the airways of smokers compared with nonsmokers (38). Later, Spira and colleagues described global alterations in gene expression between normal-appearing bronchial epithelium of healthy cancer-free smokers and that of nonsmokers (39). Importantly, irreversible changes in expression in airways of former smokers after years of smoking cessation were described that were believed to underlay the increased risk former smokers exhibit for developing lung cancer (39, 40). Alterations in the expression of microRNAs were also demonstrated between large airways of current and never-smokers (41). Notably, an 80-gene signature was derived from the transcriptome of large airway epithelial cells that can distinguish smokers without overt cancer from smokers with lung cancer despite originating from normal bronchial epithelia (42). More recently, Gustafson and coworkers derived a phosphoinositide-3-kinase (*PI3K*) pathway activation signature by using recombinant adenoviruses to express the 110α subunit of *PI3K* in primary human epithelial cells (43). The *PI3K* pathway activation signature was elevated in cytologically normal bronchial airways of smokers with lung cancer and, importantly, was decreased in the airways of high-risk smokers whose dysplastic lesions regressed after treatment with the *PI3K* inhibitor myoinositol (43). Microarray and gene expression profiling methodologies were also used to demonstrate the wide anatomical spread of the lung field cancerization. Common gene expression alterations were identified in bronchial, nasal, and buccal epithelia of smokers (44), and in a separate study, the expression of 119 genes was

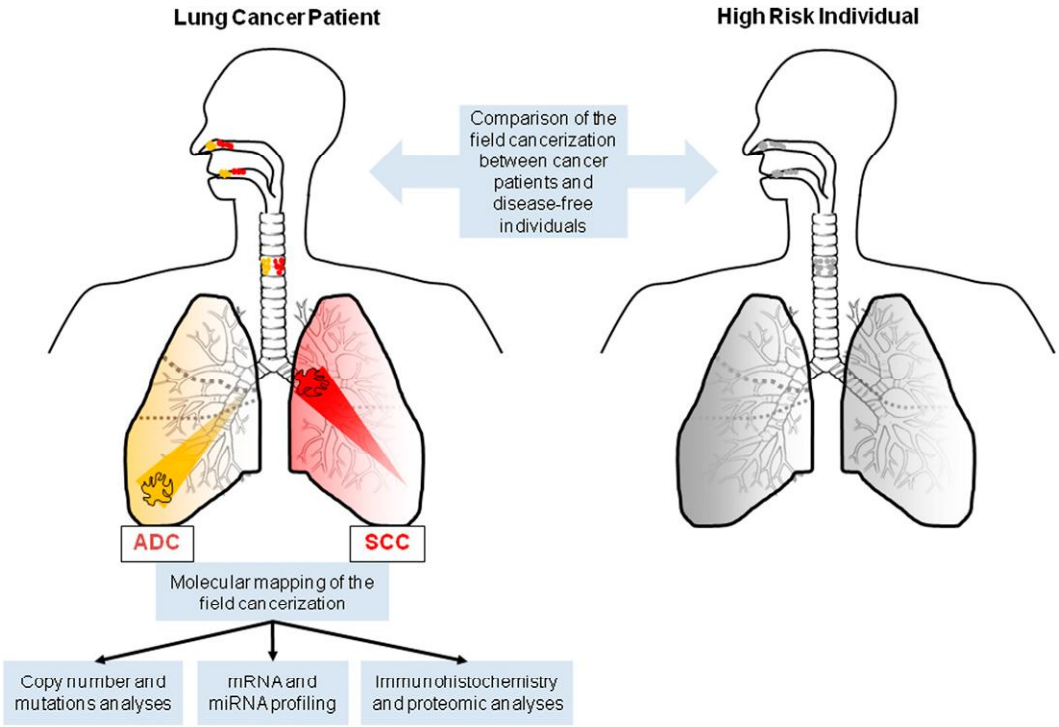


Figure 1. Molecular mapping analysis of the field cancerization in non–small cell lung cancer (NSCLC). The relevance of the lung field cancerization to the development of a particular subtype of NSCLC (i.e., adenocarcinoma compared with squamous cell carcinoma [SCC]), is still unknown, yet possible. Analyzing local and distant field of cancerization by analysis of the transcriptome of airway brushings from multiple sites independently for lung adenocarcinoma (*yellow spots*) and SCC (*red spots*) cases may shed light on events common or unique to the molecular pathogenesis of the two major subtypes of NSCLC. A "compartmental" approach coupled to a gradient or detailed molecular mapping method, which spans the tumor up to the nasal or buccal epithelium, to studying the field of cancerization may unravel biomarkers that can guide personalized prevention strategies suitable for each different NSCLC subtype. In addition, a comparison of the distant field cancerization in patients with cancer (*left*) to the expression patterns of the corresponding anatomical location in disease-free individuals (e.g., high-risk heavy smokers; *right*) would facilitate the development of efficacious markers for the detection of NSCLC. ADC = adenocarcinoma; miRNA = microRNA.

demonstrated to be affected by smoking similarly in both bronchial and nasal epithelium (45).

Field Cancerization Compartmentalization

In light of the prevalence of mutations in the *EGFR* oncogene in adenocarcinomas and in particular those occurring in never-smokers, Tang and colleagues investigated the presence of *EGFR* mutations in normal bronchial and bronchiolar epithelium adjacent to *EGFR* mutant tumors (22). *EGFR* mutations were detected in histologically normal peripheral epithelia in 44% of patients with lung adenocarcinoma with mutations but none in patients lacking mutations in the oncogene (22). Moreover, the same study highlighted more frequent *EGFR* mutations in normal epithelium within the tumor (43%) than in adjacent sites (24%), suggesting a localized field effect phenomenon for this abnormality in the respiratory epithelium of the lung (22). In addition, a higher frequency of mutations in cells obtained from small bronchi (35%) compared with bronchioles (18%) was detected (22). More recently, EGFR protein overexpression, similar to mutation of the gene, also exhibited a localized field effect, as it was more frequent in normal bronchial epithelia sites within tumors than in sites adjacent to and distant from tumors (23). Interestingly, *EGFR* copy number alteration was not evident in normal bronchial epithelia, which is in accordance with findings that *EGFR* copy number is a relatively late event in pathogenesis of adenocarcinomas (5, 23). These findings suggest that adenocarcinomas may be associated with a field cancerization dissimilar from that linked to SCCs.

The low frequency of molecular abnormalities detected in the centrally located bronchial respiratory epithelium in patients with peripheral lung adenocarcinomas, compared with specimens from patients with SCCs (28), suggests the presence of two compartments in the lung with different degrees of smoking-related genetic damage. Thus, smokers who develop SCCs display more smoking-related genetic damage in the respiratory epithelium of the central airway, whereas patients who develop adenocarcinomas exhibit molecular and histological damage mainly in the peripheral airways. Although some molecular changes (e.g., inflammation and signaling pathways activation) have been detected throughout the lung airway and include both compartments (central and peripheral airway), other aberrations have been more frequently altered in either central (e.g., LOH, genetic instability evidenced by microsatellite repeats) or peripheral (e.g., *EGFR* mutations as mentioned above) airways. These interesting observations indicate a possible compartmentalization of field cancerization and its dissimilarity between adenocarcinomas and SCCs, which may well reflect the differential mechanisms of pathogenesis of both NSCLC subtypes.

FUTURE DIRECTIONS AND CONCLUSION

Applying the same advanced high-throughput methodologies currently used in studying established tumors for the genetic analysis of lung NSCLC pre-neoplasia and histologically normal adjacent regions is expected to expand our understanding of the biology of this prevalent disease. Next-generation sequencing technology, through whole-genome, whole-exome, and whole-transcriptome approaches, holds great promise for providing invaluable insights into NSCLC biology, diagnosis, prevention, and therapy (46). An important step in this direction was a recent study in which RNA of bronchial airway epithelial cell brushings from healthy never-smokers and smokers with and without lung cancer was analyzed by RNA sequencing (47) and provided additional insight besides that provided when using microarray technology.

Earlier findings demonstrated that centrally located lung SCCs and peripherally located lung adenocarcinomas elicit and perpetuate differential effects on the airway epithelia (4). Changes in expression in the lung field of injury have been shown to be similar in the large and small airways, and it is unknown whether they are associated with the development of the particular subtype of NSCLC. Addressing this question may be highly important, because both NSCLC subtypes display different genomic features (2) and, therefore, are clinically managed by significantly dissimilar treatment strategies, let alone differences among various subtypes of lung adenocarcinomas. Moreover, revisiting the field cancerization effect using a compartmental coupled with a gradient or detailed molecular mapping approach in patients with cancer and disease-free individuals (Figure 1) will shed light on events in the early pathogenesis of lung adenocarcinomas and SCCs and unravel biomarkers that can guide targeted and personalized chemoprevention strategies suitable for each different NSCLC subtype as well as detection efforts, in particular using less invasive sites.

Despite numerous efforts that have centered on increasing our understanding of the biology of lung cancer, this malignancy still composes the biggest share of cancer-related deaths in the United States and worldwide. Compared with advances in targeted and personalized therapy of NSCLC, little progress has been made in the tailored prevention of this fatal malignancy. This may change with the recent encouraging and significant findings of the National Lung Screening Trial (11). Various molecular markers and expression classifiers previously described in the lung airways and in less-invasive sites of field cancerization (e.g., nasal epithelium) can aid in selecting high-risk individuals best suited for computed tomography screening, for example. A comprehensive analysis of early molecular events in NSCLC pathogenesis will undoubtedly unravel biomarkers that can guide future chemoprevention strategies.

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;61:69–90.
2. Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med* 2008;359:1367–1380.
3. Wistuba II, Gazdar AF. Lung cancer preneoplasia. *Annu Rev Pathol* 2006;1:331–348.
4. Wistuba II. Genetics of preneoplasia: lessons from lung cancer. *Curr Mol Med* 2007;7:3–14.
5. Yatabe Y, Borczuk AC, Powell CA. Do all lung adenocarcinomas follow a stepwise progression? *Lung Cancer* 2011;74:7–11.
6. Kabat GC, Wynder EL. Lung cancer in nonsmokers. *Cancer* 1984;53: 1214–1221.
7. Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung Cancer* 2001;31:139–148.
8. Rudin CM, Avila-Tang E, Harris CC, Herman JG, Hirsch FR, Pao W, Schwartz AG, Vahakangas KH, Samet JM. Lung cancer in never smokers: molecular profiles and therapeutic implications. *Clin Cancer Res* 2009;15:5646–5661.
9. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers–a different disease. *Nat Rev Cancer* 2007;7:778–790.
10. Gold KA, Kim ES, Lee JJ, Wistuba II, Farhangfar CJ, Hong WK. The battle to personalize lung cancer prevention through reverse migration. *Cancer Prev Res (Phila)* 2011;4:962–972.
11. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
12. Yatabe Y. EGFR mutations and the terminal respiratory unit. *Cancer Metastasis Rev* 2010;29:23–36.
13. Yatabe Y, Mitsudomi T, Takahashi T. TTF-1 expression in pulmonary adenocarcinomas. *Am J Surg Pathol* 2002;26:767–773.

14. Yuspa SH. The pathogenesis of squamous cell cancer: lessons learned from studies of skin carcinogenesis–thirty-third G. H. A. Clowes Memorial Award Lecture. *Cancer Res* 1994;54:1178–1189.
15. Aldaz CM, Conti CJ, Larcher F, Trono D, Roop DR, Chesner J, Whitehead T, Slaga TJ. Sequential development of aneuploidy, keratin modifications, and gamma-glutamyltransferase expression in mouse skin papillomas. *Cancer Res* 1988;48:3253–3257.
16. Aldaz CM, Trono D, Larcher F, Slaga TJ, Conti CJ. Sequential trisomization of chromosomes 6 and 7 in mouse skin premalignant lesions. *Mol Carcinog* 1989;2:22–26.
17. Bianchi AB, Aldaz CM, Conti CJ. Nonrandom duplication of the chromosome bearing a mutated Ha-ras-1 allele in mouse skin tumors. *Proc Natl Acad Sci USA* 1990;87:6902–6906.
18. Wistuba II, Behrens C, Milchgrub S, Bryant D, Hung J, Minna JD, Gazdar AF. Sequential molecular abnormalities are involved in the multistage development of squamous cell lung carcinoma. *Oncogene* 1999;18:643–650.
19. Mao L, Lee JS, Kurie JM, Fan YH, Lippman SM, Lee JJ, Ro JY, Broxson A, Yu R, Morice RC, *et al.* Clonal genetic alterations in the lungs of current and former smokers. *J Natl Cancer Inst* 1997;89:857–862.
20. Wistuba II, Lam S, Behrens C, Virmani AK, Fong KM, LeRiche J, Samet JM, Srivastava S, Minna JD, Gazdar AF. Molecular damage in the bronchial epithelium of current and former smokers. *J Natl Cancer Inst* 1997;89:1366–1373.
21. McCaughan F, Pole JC, Bankier AT, Konfortov BA, Carroll B, Falzon M, Rabbitts TH, George PJ, Dear PH, Rabbitts PH. Progressive 3q amplification consistently targets SOX2 in preinvasive squamous lung cancer. *Am J Respir Crit Care Med* 2010;182:83–91.
22. Tang X, Shigematsu H, Bekele BN, Roth JA, Minna JD, Hong WK, Gazdar AF, Wistuba II. EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. *Cancer Res* 2005;65:7568–7572.
23. Tang X, Varella-Garcia M, Xavier AC, Massarelli E, Ozburn N, Moran C, Wistuba II. Epidermal growth factor receptor abnormalities in the pathogenesis and progression of lung adenocarcinomas. *Cancer Prev Res (Phila)* 2008;1:192–200.
24. Noguchi M. Stepwise progression of pulmonary adenocarcinoma–clinical and molecular implications. *Cancer Metastasis Rev* 2010;29:15–21.
25. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* 1953;6:963–968.
26. Auerbach O, Stout AP, Hammond EC, Garfinkel L. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. *N Engl J Med* 1961;265:253–267.
27. Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. *Cancer Prev Res (Phila)* 2008;1:396–403.
28. Wistuba II, Behrens C, Virmani AK, Mele G, Milchgrub S, Girard L, Fondon JW III, Garner HR, McKay B, Latif F, *et al.* High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. *Cancer Res* 2000;60:1949–1960.
29. Wistuba II, Berry J, Behrens C, Maitra A, Shivapurkar N, Milchgrub S, Mackay B, Minna JD, Gazdar AF. Molecular changes in the bronchial epithelium of patients with small cell lung cancer. *Clin Cancer Res* 2000;6:2604–2610.
30. Nelson MA, Wymer J, Clements N Jr. Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers. *Cancer Lett* 1996; 103:115–121.
31. Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, Baylin SB, Herman JG. Aberrant methylation of p16 (INK4A) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc Natl Acad Sci USA* 1998;95:11891–11896.
32. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, Grimes MJ, Harms HJ, Tellez CS, Smith TM, *et al.* Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res* 2002;62:2370–2377.
33. Powell CA, Klares S, O'Connor G, Brody JS. Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer. *Clin Cancer Res* 1999;5:2025–2034.
34. Franklin WA, Gazdar AF, Haney J, Wistuba II, La Rosa FG, Kennedy T, Ritchey DM, Miller YE. Widely dispersed p53 mutation in respiratory epithelium: a novel mechanism for field carcinogenesis. *J Clin Invest* 1997;100:2133–2137.
35. Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, Seidl S, Minna JD, Gazdar AF. Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. *Int J Cancer* 2003;107:612–616.
36. Soria JC, Rodriguez M, Liu DD, Lee JJ, Hong WK, Mao L. Aberrant promoter methylation of multiple genes in bronchial brush samples from former cigarette smokers. *Cancer Res* 2002;62:351–355.
37. Heller G, Zielinski CC, Zochbauer-Muller S. Lung cancer: from single-gene methylation to methylome profiling. *Cancer Metastasis Rev* 2010;29:95–107.
38. Hackett NR, Heguy A, Harvey BG, O'Connor TP, Luettich K, Flieder DB, Kaplan R, Crystal RG. Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers. *Am J Respir Cell Mol Biol* 2003;29:331–343.
39. Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, Palma J, Brody JS. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci USA* 2004;101:10143–10148.
40. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol* 2007;8:R201.
41. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, Lu J, Liu G, Bowers J, Vaziri C, *et al.* MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proc Natl Acad Sci USA* 2009;106:2319–2324.
42. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, Gilman S, Dumas YM, Calner P, Sebastiani P, *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 2007;13:361–366.
43. Gustafson AM, Soldi R, Anderlind C, Scholand MB, Qian J, Zhang X, Cooper K, Walker D, McWilliams A, Liu G, *et al.* Airway PI3K pathway activation is an early and reversible event in lung cancer development. *Sci Transl Med* 2010;2:26ra25.
44. Sridhar S, Schembri F, Zeskind J, Shah V, Gustafson AM, Steiling K, Liu G, Dumas YM, Zhang X, Brody JS, *et al.* Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* 2008;9:259.
45. Zhang X, Sebastiani P, Liu G, Schembri F, Dumas YM, Langer EM, Alekseyev Y, O'Connor GT, Brooks DR, Lenburg ME, *et al.* Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. *Physiol Genomics* 2010;41:1–8.
46. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 2010; 11:685–696.
47. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, Luo L, Zhang XH, Xiao J, Alekseyev YO, *et al.* Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-seq. *Cancer Prev Res (Phila)* 2011;4:803–817.

G-Protein Coupled Receptor Family C, Group 5, Member A (*GPRC5A*) Expression Is Decreased in the Adjacent Field and Normal Bronchial Epithelia of Patients with Chronic Obstructive Pulmonary Disease and Non–Small-Cell Lung Cancer

Junya Fujimoto, MD, PhD, Humam Kadara, PhD,* Melinda M. Garcia, MS,* Mohamed Kabbout, PhD,* Carmen Behrens, MD,* Diane D. Liu, MS,† J. Jack Lee, DDS, PhD,† Luisa M. Solis, MD,‡ Edward S. Kim, MD,* Neda Kalhor, MD,‡ Cesar Moran, MD,‡ Amir Sharafkhaneh, MD, PhD,§ Reuben Lotan, PhD,*# and Ignacio I. Wistuba MD*‡*

Introduction: Understanding oncogenes and tumor suppressor genes expression patterns is essential for characterizing lung cancer pathogenesis. We have previously demonstrated that m*Gprc5a*/h*GPRC5A* is a lung-specific tumor suppressor evidenced by inflammation-mediated tumorigenesis in *Gprc5a*-knockout mice. The implication of *GPRC5A* in human lung cancer pathogenesis, including that associated with inflammatory chronic obstructive pulmonary disease (COPD), a risk factor for the malignancy, remains elusive.

Methods: We sought to examine *GPRC5A* immunohistochemical expression in histologically normal bronchial epithelia (NBE) from lung disease-free never- and ever-smokers (n = 13 and n = 18, respectively), from COPD patients with (n = 26) and without cancer (n = 24) and in non-small cell lung cancers (NSCLCs) (n = 474). Quantitative assessment of GPRC5A transcript expression in airways (n = 6), adjacent NBEs (n = 29) and corresponding tumors (n = 6) from 6 NSCLC patients was also performed.

Results: GPRC5A immunohistochemical expression was significantly lower in tumors compared to uninvolved NBE ($p < 0.0001$) and was positively associated with adenocarcinoma histology ($p < 0.001$). GPRC5A airway expression was highest in lung disease-free NBE, decreased and intermediate in NBE of cancer-free COPD patients ($p = 0.004$) and further attenuated and lowest in epithelia of COPD patients with adenocarcinoma and SCC ($p < 0.0001$). Furthermore, *GPRC5A* mRNA was significantly decreased in NSCLCs and corresponding NBE compared to uninvolved normal lung ($p = 0.03$).

Conclusions: Our findings highlight decreased *GPRC5A* expression in the field cancerization of NSCLC, including that associated with lung inflammation. Assessment of the use of *GPRC5A* expression as a risk factor for NSCLC development in COPD patients is warranted.

KEY WORDS: Field cancerization, Chronic obstructive pulmonary disease, Non–small-cell lung cancer, g-protein coupled receptor family C, group 5, member A, gene expression.

(*J Thorac Oncol.* 2012;7: 1747–1754)

Lung cancer, the majority of which is non–small cell lung cancer (NSCLC), is the leading cause of deaths in the United States and worldwide.[1] The high mortality associated with lung cancer is in part because of late diagnosis after regional or distant spread of the disease.[2] Improved clinical management of NSCLC is tightly linked to the identification of new effective early biomarkers that can spear novel strategies for early detection, prevention, and treatment.[2,3]

The majority (85%) of diagnosed NSCLC cases are attributable to cigarette smoking.[4,5] Auerbach et al. [6] earlier showed that tobacco carcinogen exposure causes multifocal and clonal histopathological changes in the airway epithelia of smokers, suggestive of a field cancerization in the lung. In addition, cigarette smoking perpetuates inflammation throughout the smoking-exposed airway epithelia of heavy smokers,[7] which was suggested, to lead to onset of lung cancer development.[8] Chronic obstructive pulmonary disease (COPD) of the lung is an inflammatory condition that is, like lung cancer, causally linked to cigarette smoking[9,10] and is a major cause of mortality in the United States.[11] Moreover, preinvasive lung cancer lesions are common (approximately 50%) in airways of COPD patients.[12] Importantly, although phenotypically healthy smokers comprise a significant population at risk for

Departments of *Thoracic/Head and Neck Medical Oncology, †Biostatistics and Applied Mathematics, ‡Pathology, University of Texas M.D. Anderson Cancer Center, Houston, Texas; and §Department of Medicine, Section of Pulmonary, Critical Care and Sleep Medicine, Baylor College of Medicine, Houston, Texas.

#Deceased.

J. Fujimoto and H. Kadara contributed equally to this work and should be considered first authors.

Disclosure: The authors declare no conflict of interest.

Address for correspondence: Junya Fujimoto, MD, PhD, Department of Thoracic/Head and Neck Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX. E-mail: jfujimot@mdanderson.org

ISSN: 1556-0864/12/0712-1747

lung cancer,[5,7] COPD increases lung cancer risk by 4.5-fold, more than either age or quantity of smoking.[11,13,14] In addition, 33% of patients with mild-to-moderate COPD eventually die because of lung cancer burden.[11] Although smoking is the main cause of both diseases, only 10% to 15% of smokers develop COPD and/or lung cancer, suggesting that other factors, including genetic variation and expression, may differ across individuals in response to cigarette smoke.[4,5,9,10] It is plausible to assume that understanding common expression patterns between both diseases for subsequent identification of biological markers that explain risk of disease onset and/or progression will favorably promote the clinical management of COPD-associated lung cancer.[11]

Our group has previously demonstrated that loss of G-protein coupled receptor, family C, group 5, member A (m*Gprc5a*/h*GPRC5A*), exemplified by the *Gprc5a*-knockout mouse model, leads to chronic inflammation and spontaneous lung tumor development.[15,16] However, the expression of this tumor suppressor in human lung NSCLCs, including those associated with inflammatory conditions such as COPD, is unknown. We investigated the immunohistochemical expression of *GPRC5A* in a large series of NSCLC histological tissue specimens and in normal bronchial epithelia (NBE) from lung-disease–free individuals, COPD patients and from COPD patients with lung cancer and its transcript expression in independent resected normal epithelia and corresponding tumors. Our findings reveal marked reduced GPRC5A expression in human NSCLC and in the histologically normal field cancerization associated with COPD and lung cancer. Our study pinpoints to a potential role of this tumor suppressor in the progression of COPD-associated NSCLC, which warrants further studies to assess its use as a risk marker for this disease.

MATERIALS AND METHODS

NSCLC Tissue mMcroarrays and Normal Bronchial Epithelial Specimens

Tissue microarrays (TMAs) used in this study comprised 474 surgically resected NSCLC tumor specimens (308 adenocarcinomas, 166 squamous cell carcinomas [SCCs]) collected under an Institutional Review Board protocol and archived as formalin-fixed paraffin-embedded (FFPE) specimens in the University of Texas Specialized Program of Research Excellence thoracic tissue bank at the University of Texas M.D. Anderson Cancer Center. Clinicopathological features of the NSCLC patients examined are summarized in Table 1. TMAs were prepared with a manual tissue arrayer (Advanced Tissue Arrayer ATA100, Chemicon International, Temecula, CA) using 1-mm diameter cores in triplicate for tumors, as described previously.[17] Histological sections 4 μm in thickness were then prepared for subsequent immunohistochemistry analysis. FFPE specimens of NBE from 50 patients with COPD with forced expiratory volume in 1 second/forced vital capacity ratio of 70% or less, a smoking history of 30 or more pack-years, and collected at Baylor college of Medicine (Houston, TX) were included in the study, including 24 cancer-free cases and 26 cases with NSCLC. In addition, NBE specimens from never smokers (n = 13) and ever smokers (n = 18) without lung cancer and with no or mild airway obstruction status were analyzed.

TABLE 1. Clinicopathological Features of NSCLC Patients in Tissue Microarray Sets Used in the Study

Covariate	Levels	n (%)
Sex	Female	240 (50.6)
	Male	234 (49.4)
Histology	Adenocarcinoma	308 (65.0)
	Squamous cell carcinoma	166 (35.0)
Stage	I	302 (63.7)
	II	86 (18.1)
	III or IV	86 (18.1)
Grade*	Well	36 (10.4)
	Moderate	193 (55.9)
	Poor	116 (33.6)
Smoking status	Current	178 (37.6)
	Former	229 (48.3)
	Never	67 (14.1)
Tobacco history	No	67 (14.1)
	Yes	407 (85.9)

*Information on differentiation grade was available in a subset of NSCLCs analyzed.
NSCLC, non–small-cell lung cancer.

Immunohistochemical Analysis

Polyclonal antirabbit antibodies raised against human GPRC5A were generated by peptide synthesis and site-directed carrier conjugation using keyhole limpet hemocyanin as a custom service by Zymed Laboratories Inc. (South San Francisco, CA). The synthetic peptide (cysteine)-PSPYKDYEVKKEGS-COOH, corresponding to amino acids 344–357 in the human GPRC5A C-terminus, was covalently linked to keyhole limpet hemocyanin via an added cysteine residue, and polyclonal antipeptide antibodies were prepared in rabbits. Sera were confirmed to contain high-titer antibodies against the specific peptide using enzyme-linked immunosorbent assays. TMAs and the histologic sections (4 μm thick) of surgical resected specimens were deparaffinized and hydrated, and antigen retrieval was performed using a decloaker with Dako target retrieval system at a pH of 6.0 (Dako North America, Inc., Carpinteria, CA). Intrinsic peroxidase activity was blocked by 3% methanol and hydrogen peroxide for 12 minutes and serum-free protein block (Dako) was used for 30 minutes for blocking nonspecific antibody binding. Slides were then incubated with the antibodies against human GPRC5A (1:200 dilution) at 4°C overnight. After three washes in Tris-buffered saline, slides were incubated for 30 minutes with Dako Envision + Dual Link at room temperature. After three additional washes, slides were incubated with Dako chromogen substrate for 5 minutes and were counterstained with hemotoxylin for another 5 minutes. FFPE whole-section specimens, except for the omission of the primary antibodies, were used as negative controls. The intensity and extent of cytoplasmic and nuclear GPRC5A immunostaining were evaluated using a light microscope (magnification, × 20)

independently by two pathologists (JF and IIW). GPRC5A immunoreactivity was mainly cytoplasmic, which was quantified using a four-value intensity score (0, none; 1, weak; 2, moderate; and 3, strong) and the percentage (0–100%) of the extent of reactivity. A final cytoplasmic expression score was obtained by multiplying the intensity and reactivity extension values (range, 0–300).

GPRC5A Transcript Expression in Resected NSCLC Specimens and Adjacent Airway Epithelia

GPRC5A mRNA expression was assessed in ever-smoker NSCLCs and their uninvolved normal lung parenchyma tissues as well as in NBE collected by brushing of multiple corresponding airways from the same lobectomy- and pneumonectomy-resected specimens. Tumor tissues from six ever-smoker NSCLCs and normal lung specimens ($n = 6$) as well as NBE by brushings ($n = 29$) from the same patients were obtained under an Institutional Review Board approved protocol in which specimens were collected from patients who had signed letters of consent. NBE were collected from multiple adjacent to tumor airways by brushing each site independently using Cytosoft cytology brushes (Cardinal Health, Dublin, OH). Parallel brushes were used for touch-prep for cytological assessment by pan-cytokeratin staining, which revealed epithelial content more than 90%. Normal histology was determined by hematoxylin and eosin staining. Brushes were immediately placed in Qiazol lysis reagent (Qiagen, Valencia, CA) and stored in −80°C until further processing. NSCLC and normal lung specimens were shaved for histological assessment of percentage of tumor content and malignant cells and for corresponding RNA isolation. Total RNA from all samples was purified using the miRNeasy kit (Qiagen) according to the manufacturer's instructions. RNA quantity was determined using the NanoDrop spectrophotometer, (Thermo Scientific, Wilmington, DE), and quality was determined by analysis of RNA integrity with Agilent Bioanalyzer 2000 (Agilent Technologies, Santa Clara, CA).

Quantitative-Real Time Polymerase Chain Reaction

Total RNA (150 ng) was reverse-transcribed using the high-capacity RNA-to-cDNA kit (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Quantitative real-time polymerase chain reaction (PCR) was performed using TaqMan® (Applied Biosystems, Foster City, CA) gene expression assays for *GPRC5A* (Hs00173681_m1) and beta-actin (*ACTB*) (Hs99999903_m1) primers (Life Technologies) on a 7900HT Fast-Real-Time PCR System (Life Technologies) according to the manufacturer's instructions. All samples were done in triplicates and normalized to *ACTB*. Relative quantification was calculated using the comparative cycle threshold method as previously described.[18]

Statistical Analysis

Summary statistics, including frequency tabulation, means, standard deviations, median, and range, were given to describe patient characteristics. Wilcoxon rank sum test or Kruskal–Wallis test was used to compare GPRC5A immunohistochemical expression between two levels or among more than two levels, when appropriate. General linear model was applied to test the difference of GPRC5A expression across NBE from COPD- and cancer-free never and ever smokers, cancer-free COPD patients, as well as NBE from patients with both COPD and cancer. Repeated measures analysis was used to determine significance of *GPRC5A* transcript variation across matched NSCLCs, airways, and uninvolved normal lung. All statistical tests were two-sided, and p values of 0.05 or less were considered to be statistically significant. Statistical analysis was performed with standard statistical software, including SAS Release 9.1.3 for Windows and S-Plus 8.0 for Windows.

RESULTS

GPRC5A Immunohistochemical Expression in Human NSCLC

We have previously demonstrated that m*Gprc5a*/h*GPRC5A* is a lung-specific tumor suppressor evidenced by spontaneous tumor development in mice with knockout of this gene as well as loss of the transformed phenotype in malignant NSCLC cell lines after *GPRC5A* overexpression.[16] We sought to examine the immunohistochemical expression patterns of GPRC5A tumor suppressor in NSCLC, which is unknown, and its association with clinicopathological features of the disease, including the two major histologic types examined, adenocarcinoma and SCC. The clinicopathological characteristics of the NSCLC patients from which FFPE tumor specimens were analyzed are detailed in Table 1. Representative photomicrographs of GPRC5A immunohistochemical expression in human lung adenocarcinoma (left) and SCC (right) are depicted in Figure 1*A*. GPRC5A immunoreactivity was mainly cytoplasmic. GPRC5A expression was significantly lower in the adenocarcinomas ($n = 308$; mean, 48.91 ± 36; median, 41.67; minimum, 0; maximum, 180) and SCCs ($n = 166$; mean, 23.47 ± 22.79; median, 18.33; minimum, 0; maximum, 110) examined compared with the uninvolved normal bronchial epithelia (mean, 164.52 ± 34.04; median, 160; minimum, 100; maximum, 230) from lung-disease–free smokers and never smokers ($p < 0.0001$) (Fig. 1*B*). Moreover, GPRC5A immunohistochemical expression was significantly lower in SCCs compared with that in adenocarcinomas ($p < 0.0001$) (Fig. 1*B*). We then correlated GPRC5A with other clinicopathological features for all NSCLCs. Notably, GPRC5A expression was significantly increased in NSCLC tumor specimens from never smokers ($n = 67$) (mean, 53.94 ± 39.94; median, 40; minimum, 0; maximum, 163.33) compared with ever smokers ($n = 407$) (mean, 37.71 ± 32.65; median, 30; minimum, 0; maximum, 180) ($p = 0.001$) (Fig. 1*C*). GPRC5A expression was also significantly different among never, former and current smoker NSCLCs ($p = 0.003$) (Fig. 1*D*). In addition, GPRC5A expression was significantly positively associated with well-differentiated tumor grade in a subset of NSCLC tumors ($n = 346$) examined with available grade information ($p = 0.004$) (Fig. 1*E*). It is important to note that when we examined each histology type separately, there were no statistically significant

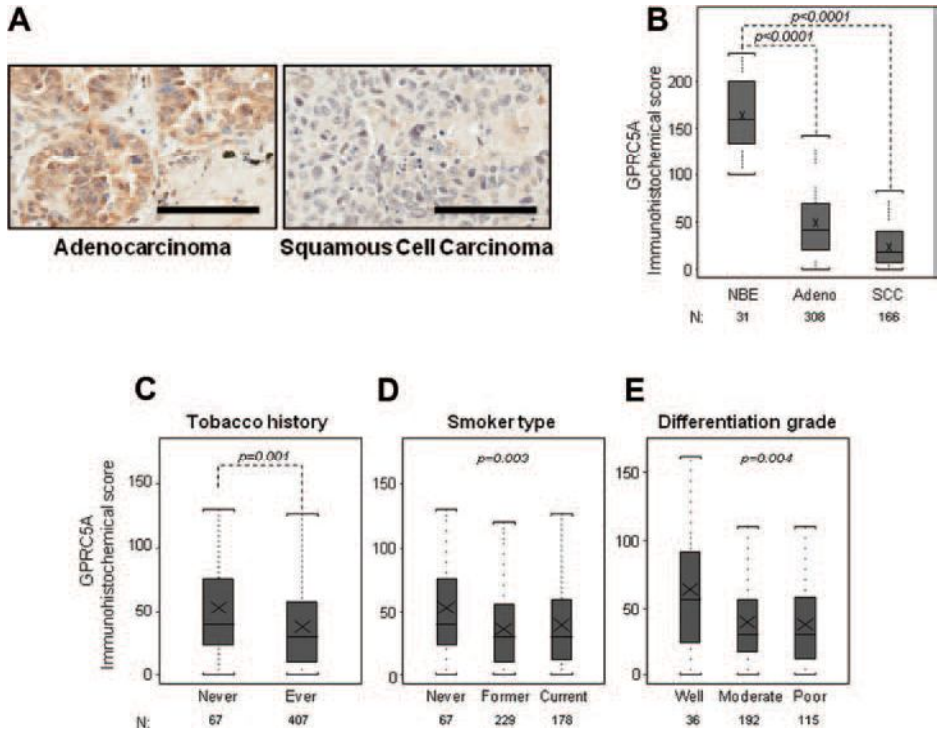


FIGURE 1. GPRC5A immunohistochemical expression in NSCLC histological tissue specimens. *A*, Representative photomicrographs of GPRC5A immunohistochemical expression in lung adenocarcinoma (left) and SCC (right) observed under a light microscope at a magnification of 20 ×. Bars, 100 μm. *B*, Box-plots depicting differences in GPRC5A immunohistochemical expression among NBE, Adeno, and SCCs. *C*, Box-plots depicting GPRC5A expression based on tobacco history (never versus ever smokers), (*D*), smoking status (never, former and current), and (*E*), differentiation grade (well, moderate, and poor). *p* values based on the Wilcoxon rank sum test and Kruskal–Wallis test. GPRC5A, G-protein coupled receptor family C, group 5, member A; NSCLC, non–small-cell lung cancer; SCC, squamous cell carcinomas; NBE, normal bronchial epithelia; Adeno, adenocarcinomas.

correlations between GPRC5A expression and patient's smoking history and status, and tumor-differentiation grade.

GPRC5A Immunohistochemical Expression in NBE from COPD Patients

We have previously shown that loss of m*Gprc5a*/h*GPRC5A* tumor suppressor leads to chronic-inflammation–mediated prosurvival signaling and transformation of nonmalignant lung epithelial cells.[15] The relevance of GPRC5A expression to human NSCLC pathogenesis, including that associated with COPD, a chronic inflammatory condition and risk factor for lung cancer, remains elusive. In light of the strong association between inflammation, COPD and NSCLC,[10,11] we sought to examine the expression of GPRC5A NBE in COPD-associated adenocarcinoma and SCC. GPRC5A expression was assessed in histological tissue specimens of NSCLC and NBE, including those from lung cancer-free COPD and from NSCLC patients. Representative photomicrographs of GPRC5A immunohistochemical staining in the different NBE are depicted in Figure 2*A*. GPRC5A immunoreactivity seemed to be highest in NBE from lung-disease–free never and ever smokers, successively significantly decreased in NBE of COPD patients who are cancer-free, and lowest in NBE of patients with both COPD and adenocarcinoma or SCC (Fig. 2*B*). A general linear model demonstrated a significant gradual decrease of GPRC5A expression from NBE of disease-free never smokers to NBE from patients with COPD and adenocarcinoma or SCC ($p < 0.0001$) (Fig. 2*B*). The mean (177.31 ± 30.66) and median (195; minimum, 130; maximum, 220) of GPRC5A expression in NBE of lung-disease–free never smokers were higher than that of smokers (mean, 155.28 ± 34.15, median, 147.5; minimum, 100; maximum, 230) although the differences were not statistically significant (Fig. 2*B*). Moreover, GPRC5A expression was significantly higher in NBE of lung-disease–free smokers relative to NBE in cancer-free COPD patients (mean, 85.48 ± 53.55, median, 72.88; minimum, 20; maximum, 211) ($p = 0.007$), which in turn was significantly higher compared with the expression in NBE from COPD patients with adenocarcinoma (mean, 48.15 ± 52.84, median, 35; minimum, 0; maximum, 170) ($p < 0.0001$) or SCC (mean, 32.24 ± 34.9, median, 12.5; minimum, 0; maximum, 101.69) ($p < 0.0001$) (Fig. 2*B*). Furthermore, although GPRC5A expression was lower in NBE of COPD patients with SCC compared with those with COPD and adenocarcinoma, the difference was not statistically significant (Fig. 2*B*). These findings suggest that reduced expression of the GPRC5A tumor suppressor may be implicated in the pathogenesis of NSCLC associated with inflammatory COPD.

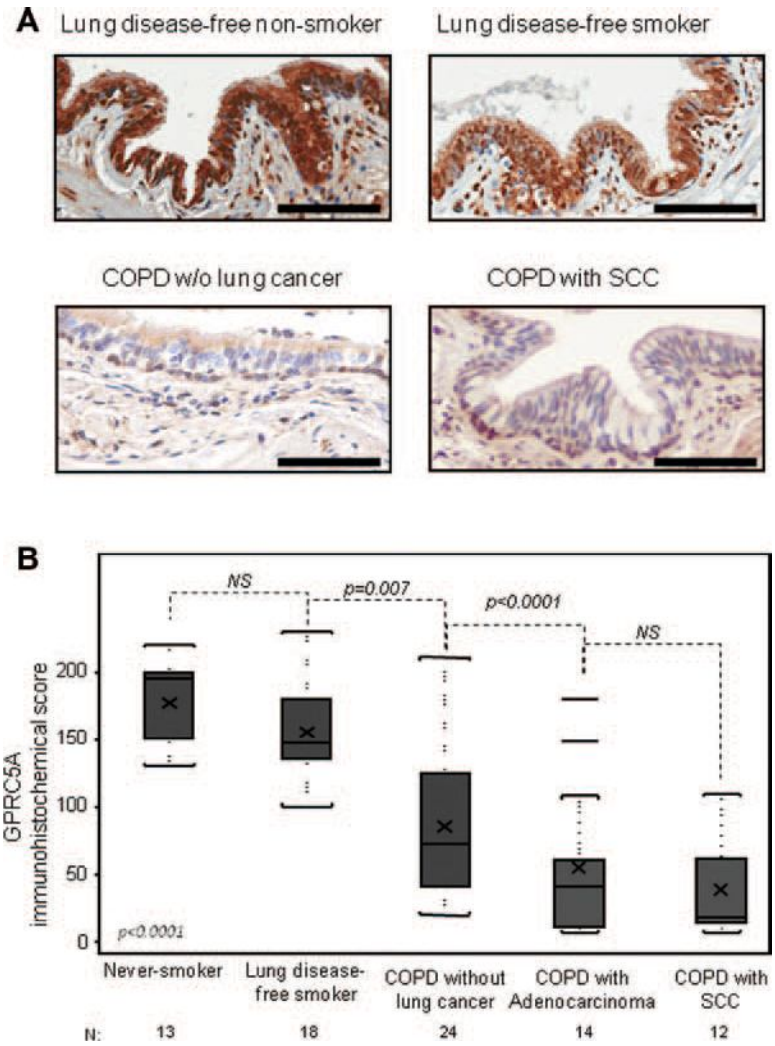


FIGURE 2. Reduced GPRC5A expression in the sequential pathogenesis of COPD-associated NSCLC. *A*, Representative photomicrographs of GPRC5A immunohistochemical expression in NBE from disease-free never and ever smokers (upper left and upper right, respectively) and in NBE from cancer-free COPD patients (bottom left) and NSCLC patients (bottom right) with COPD observed under a light microscope at a magnification of × 20. Bars, 100 μm. *B*, Box-plots depicting quantification of GPRC5A immunohistochemical expression represented in (*A*). Main *p* value signifies statistical significance of reduced GPRC5A expression in the sequential pathogenesis of COPD-associated NSCLC and among the five groups by the general linear model and Kruskal–Wallis test. *p* values signifying the indicated pair wise comparisons among the groups were obtained by the Wilcoxon rank sum test. GPRC5A, G-protein coupled receptor family C, group 5, member A; NBE, normal bronchial epithelia; NSCLC, non–small-cell lung cancer; COPD, chronic obstructive pulmonary disease; SCC, squamous cell carcinomas; NS, not significant.

We then examined GPRC5A expression in epithelial cells from different compartments of the lung (bronchial, bronchiolar and alveolar epithelia and cells). GPRC5A expression was statistically significantly lower in bronchi and bronchioles from cancer-free COPD patients ($p < 0.005$) and patients with both COPD and SCCs or adenocarcinomas ($p < 0.0001$) compared with lung-disease–free never smokers (Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/JTO/A334). In contrast, GPRC5A expression in alveolar compartment remained high (score average 172.2–261.5) in all groups of cases examined, and there were no statistically significant differences based on COPD status (Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/JTO/A334).

GPRC5A Expression in Smoker NSCLC Field Cancerization

Our previous findings on *GPRC5A* tumor-suppressive function[15,16,19] as well as our current observation on significant and progressive reduced expression of this tumor suppressor in NBE from lung cancer-free COPD patients and patients with NSCLC, prompted us to probe whether *GPRC5A* expression exhibits a field cancerization pattern in NSCLC. We sought

to assess *GPRC5A* transcript expression in multiple bronchial sites ($n = 29$) as well as in the tumors ($n = 6$) and uninvolved normal lung parenchyma ($n = 6$) in six NSCLC cases (3 SCCs and 3 adenocarcinomas). Quantitative real-time PCR analysis demonstrated progressively reduced *GPRC5A* expression from normal uninvolved lung parenchyma (relative quantification [RQ] mean, 4.18 ± 0.62; median, 4.16; minimum, 3.51; maximum, 5.04) to histologically normal and adjacent bronchial epithelia (RQ mean, 3.12 ± 1.94; median, 2.86,; minimum, 0.59; maximum, 10.12) to NSCLC tumors (RQ mean, 1.35 ± 0.71; median 1.31; minimum 0.24; maximum 2.30), which exhibited on an average the lowest expression of the gene (Fig. 3*A* and *B*) ($p = 0.03$). These findings suggest that *GPRC5A* expression is reduced in the smoking-injured field cancerization in NSCLC highlighting a potentially strong implication of this tumor suppressor in the pathogenesis of this malignancy.

DISCUSSION

In this study, we assessed the expression of *GPRC5A*, which we had found to exhibit lung-specific tumor-suppressor properties in mice,[15,16,19] in human NSCLC histological tissue specimens, as well as in NBE from COPD and lung adenocarcinoma and SCC patients. We found that the protein product of human *GPRC5A* was lower in lung tumors compared with uninvolved histologically NBE, and was significantly associated with adenocarcinoma histology. *GPRC5A* airway expression seemed to exhibit a progressive decrease in the field cancerization of COPD-related NSCLC with highest expression in bronchial epithelia obtained from disease-free individuals, intermediate immunoreactivity in normal epithelia from cancer-free COPD patients, and lowest in normal epithelia from patients with both COPD and adenocarcinoma and SCC histologies. Furthermore, *GPRC5A* transcript expression was also lower in histologically NBE as well as in adjacent and corresponding smoker NSCLC tumors, irrespective of COPD status, compared with uninvolved normal lung suggestive of a field cancerization-mediated expression modulation of this G-protein coupled receptor. In light of the relevance of the field cancerization phenomenon to NSCLC pathogenesis,[5–7,20,21] our findings pinpoint to a potential tumor-suppressive role, similar to that established in mice, of *GPRC5A* in the sequential development of human NSCLC, in particular those associated with inflammatory chronic obstructive disease. NSCLC and COPD are both mainly attributable to cigarette smoke[9,10,13] and are leading causes of deaths in the United States and worldwide.[1,2,11] Given that COPD is a major risk factor for lung cancer and shares various pathogenic features with lung tumors,[9,10,22–25] better molecular markers are needed to identify which COPD patients will continue to develop lung malignancies.[11] Thus, the results herein raise the intriguing possibility that *GPRC5A* loss may be a useful biomarker in assessing the risk of NSCLC development in COPD patients.

COPD and infections as well as inflammatory disorders of the respiratory tract may be linked to an increased risk of lung cancer.[26,27] We have previously demonstrated significantly increased macrophage infiltration into lungs of *Gprc5a*-knockout mice along with their direct association with adenocarcinomas, which was accompanied by higher constitutive levels of proinflammatory cytokines and chemokines and increased susceptibility to stimulation of nuclear factor-kappa B (NF-κB) activation in vivo.[15] We also showed that loss of *Gprc5a*-mediated activation of NF-κB was causally linked to macrophage recruitment and enhanced inflammation toward the creation of a tumor-promoting microenvironment.[15] It is noteworthy that lungs of patients with severe COPD exhibit macrophage/$CD8^+$ T cell and neutrophil infiltration based on histopathological studies and bronchial-biopsies.[28] Moreover, it is worthwhile to mention that we attempted to statistically assess the association between expression of GPRC5A and that of NF-κB (nuclear p65 immunoreactivity) in NSCLCs and found a significant

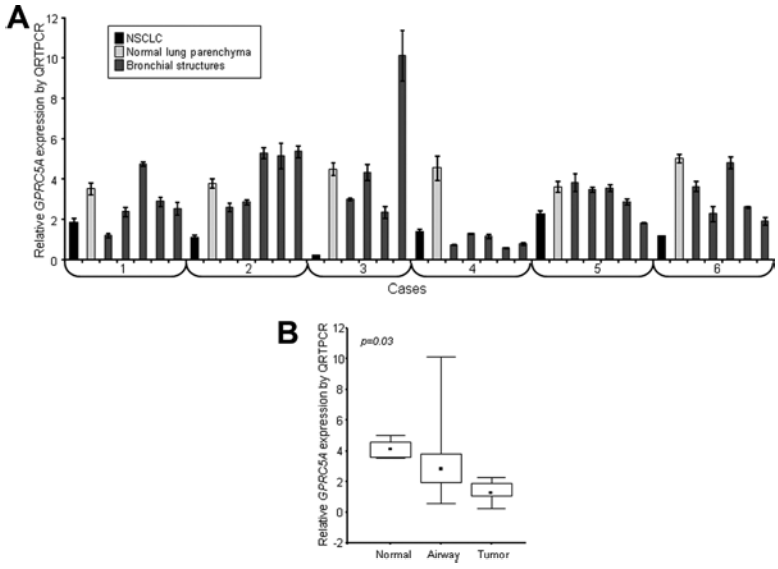


FIGURE 3. Decreased *GPRC5A* mRNA expression in histologically NBE and corresponding NSCLC tumors from resected specimens compared with matched uninvolved normal lung. Total RNA was isolated from brushings of NBE and frozen sections of matched NSCLC (T) and uninvolved NL parenchyma obtained from resected lobectomy or pneumonectomy specimens performed on smoker early-stage patients. *A*, *GPRC5A* expression was analyzed in all samples (in triplicates) by QRTPCR and normalized against that of *ACTB* to yield a relative expression by the $2^{-\Delta\Delta CT}$ method depicted in the graphs. NSCLC tumors (squamous cell carcinoma, cases 1–3, and adenocarcinomas, cases 4–6), uninvolved NL parenchyma and bronchial structures are labeled by the indicated colors. *B*, *GPRC5A* transcript expression among uninvolved NL tissue, NBE, and NSCLC tumors was assessed statistically by repeated measures analysis. NBE, normal bronchial epithelia; NSCLC, non–small-cell lung cancer; NL, normal lung; CT, comparative cycle threshold method.; QRTPCR, quantitative real-time polymerase chain reaction.

inverse correlation between both proteins in particular in adenocarcinomas (data not shown). In addition, it was recently shown that the incidence of hyperplastic lesions in lungs of *Gprc5a*-knockout mice was increased after exposure to nontypeable *Haemophilus* influenzae.[29] Notably, bacterial colonization, particularly with nontypeable *Haemophilus* influenzae, has been implicated as a cause of airway inflammation in COPD besides cigarette smoke.[30] In this context, the commonalities between inflammatory-mediated and histopathological mechanisms in *Gprc5a*-knockout mice and those evident in lungs of COPD patients along with our current findings on reduced *GPRC5A* expression in normal cells from COPD patients pinpoint to a tumor-suppressive role of this gene in COPD-associated human lung tumorigenesis.

We previously cloned the retinoid-regulated m*Gprc5a*/h*GPRC5A* and found that it was preferentially expressed in fetal and adult mouse and human lung tissue compared with normal specimens from other anatomically distinct organs.[31] The functional relevance of this lung-specific expression was highlighted by our earlier study demonstrating spontaneous development of adenomas and adenocarcinomas in *Gprc5a*-nockout mice, which was not observed in wild-type littermates. In addition, spontaneous tumor development in *Gprc5a*-knockout mice was characterized by late tumor onset (12–16-month-old mice) and low multiplicity.[16] Notably, we reported that exposure to the tobacco-specific carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (nicotine-derived nitrosamine ketone) augmented tumorigenesis evidenced by 6-month earlier development of lesions, two- to fourfold increased tumor incidence and multiplicity, as well as a dramatic increase in lesion size.[19] It is worthy to note that, in this study, we found significantly reduced expression of *GPRC5A* in human smoking-injured histologically NBE adjacent to lung tumors and in the localized field cancerization of smoker NSCLC patients, compared with uninvolved normal lung. Moreover, reduced *GPRC5A* expression was concomitant among airways and corresponding NSCLC compared with matched normal lung tissue. It is plausible to assume that smoking-mediated reduced airway expression of *GPRC5A* in patients may explain our previous observations on augmented tumorigenesis in *Gprc5a*-knockout mice after exposure to tobacco carcinogens. Because COPD is causally linked to smoking,[9–11,22,23] it cannot be neglected that reduced *GPRC5A* expression in NBE from COPD patients may be because of the molecular impact of smoking on the airway transcriptome. However, it is important to mention that we found significantly decreased *GPRC5A* expression in normal epithelia from COPD patients compared with epithelia obtained from disease-free smokers. Moreover, GPRC5A expression was further decreased and lowest in normal airway epithelia from COPD patients with NSCLC compared with those from cancer-free COPD patients. Thus in this context, our findings raise the possibility that *GPRC5A* expression may be impacted by cigarette smoking and may be both a cause and consequence of increased inflammation in tumor-promoting microenvironment in the lung as well as NSCLC development, which may be better delineated in future warranted studies.

Our group has previously shown that human lung and head and neck cancer cell lines exhibit reduced *GPRC5A* expression compared with their normal counterparts.[31] It is worthwhile to mention that treatment of cancer cells with all-trans retinoic acid restored GPRC5A expression in cells devoid of the tumor suppressor in part by a retinoic acid receptor-dependent manner.[31,32] In addition and in a separate study, we have previously noted reduced retinoic acid receptor beta (RARβ) in approximately 50% of adenocarcinomas and SCCs as well as loss of RARγ and retinoid X receptorβ in a significant fraction of the tumors.[33] Moreover, RARβ expression is decreased in bronchial epithelia and preneoplastic lesions representing the sequence of lung SCC pathogenesis.[34] More recently, we demonstrated, using gene expression profiling, that loss of murine Gprc5a in lung epithelial cells reduced expression of markers of squamous differentiation concomitantly with an increase in mediators of the inflammatory process (e.g., NF-κB signaling). In light of the biological connection between retinoid signaling and GPRC5A expression as well as our current findings in this revised article on the significant reduced expression of the tumor suppressor in NBE of COPD patients including those with NSCLC, it is plausible to suggest that retinoid signaling may be aberrantly regulated in pathogenesis of COPD-associated NSCLC. It is important to note that retinoids are currently being tested for their lung regenerative properties and effects in patients with COPD.[35]

In conclusion, we report herein decreased expression of *GPRC5A* in human NSCLC tissue compared with uninvolved NBE and in the field cancerization of smoker NSCLCs pinpointing to a tumor-suppressor role of this G-protein coupled receptor in the pathogenesis of this leading cause of cancer-related deaths. Moreover, our findings highlight a progressive decrease in *GPRC5A* expression in the sequential pathogenesis of NSCLCs arising in COPD patients, warranting future studies, including the analysis of lung cancer preneoplastic lesions, to assess the potential of the utility of this gene as a biomarker for lung cancer risk in COPD patients.

ACKNOWLEDGMENTS

This study was supported in part by grants from the Samuel Waxman Cancer Research Foundation and Department of Defense (DoD) grants W81XWH-04-1-0142 and W81XWH-10-1-1007 (to Ignacio I. Wistuba, MD), and the National Institute of Health, Specialized Program of Research Excellence (SPORE) in lung cancer P50CA70907 (to Ignacio I. Wistuba), and the Cancer Center Support Grant (CCSG) (CA-16672).

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012;62:10–29.
2. Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med* 2008;359:1367–1380.
3. Goldstraw P, Ball D, Jett JR, et al. Non-small-cell lung cancer. *Lancet* 2011;378:1727–1740.
4. Gazdar AF, Thun MJ. Lung cancer, smoke exposure, and sex. *J Clin Oncol* 2007;25:469–471.
5. Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. *Cancer Prev Res (Phila)* 2008;1:396–403.
6. Auerbach O, Stout AP, Hammond EC, Garfinkel L. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. *N Engl J Med* 1961;265:253–267.

7. Wistuba II, Gazdar AF. Lung cancer preneoplasia. *Annu Rev Pathol* 2006;1:331–348.
8. David H. Rudolf Virchow and modern aspects of tumor pathology. *Pathol Res Pract* 1988;183:356–364.
9. Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Br Med J* 1977;1:1645–1648.
10. Mannino DM. COPD: epidemiology, prevalence, morbidity and mortality, and disease heterogeneity. *Chest* 2002;121(5 Suppl):121S–126S.
11. Punturieri A, Szabo E, Croxton TL, Shapiro SD, Dubinett SM. Lung cancer and chronic obstructive pulmonary disease: needs and opportunities for integrated research. *J Natl Cancer Inst* 2009;101:554–559.
12. Lam S, leRiche JC, Zheng Y, et al. Sex-related differences in bronchial epithelial changes associated with tobacco smoking. *J Natl Cancer Inst* 1999;91:691–696.
13. Mannino DM, Aguayo SM, Petty TL, Redd SC. Low lung function and incident lung cancer in the United States: data From the First National Health and Nutrition Examination Survey follow-up. *Arch Intern Med* 2003;163:1475–1480.
14. Tockman MS, Anthonisen NR, Wright EC, Donithan MG. Airways obstruction and the risk for lung cancer. *Ann Intern Med* 1987;106:512–518.
15. Deng J, Fujimoto J, Ye XF, et al. Knockout of the tumor suppressor gene Gprc5a in mice leads to NF-kappaB activation in airway epithelium and promotes lung inflammation and tumorigenesis. *Cancer Prev Res (Phila)* 2010;3:424–437.
16. Tao Q, Fujimoto J, Men T, et al. Identification of the retinoic acid-inducible Gprc5a as a new lung tumor suppressor gene. *J Natl Cancer Inst* 2007;99:1668–1682.
17. Sun M, Behrens C, Feng L, et al. HER family receptor abnormalities in lung cancer brain metastases and corresponding primary tumors. *Clin Cancer Res* 2009;15:4829–4837.
18. Tahara E, Kadara H, Lacroix L, Lotan D, Lotan R. Activation of protein kinase C by phorbol 12-myristate 13-acetate suppresses the growth of lung cancer cells through KLF6 induction. *Cancer Biol Ther* 2009;8:801–807.
19. Fujimoto J, Kadara H, Men T, van Pelt C, Lotan D, Lotan R. Comparative functional genomics analysis of NNK tobacco-carcinogen induced lung adenocarcinoma development in Gprc5a-knockout mice. *PLoS ONE* 2010;5:e11847.
20. Braakhuis BJ, Tabor MP, Kummer JA, Leemans CR, Brakenhoff RH. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. *Cancer Res* 2003;63:1727–1730.
21. Gold KA, Kim ES, Lee JJ, Wistuba II, Farhangfar CJ, Hong WK. The BATTLE to personalize lung cancer prevention through reverse migration. *Cancer Prev Res (Phila)* 2011;4:962–972.
22. Cohen BH, Diamond EL, Graves CG, et al. A common familial component in lung cancer and chronic obstructive pulmonary disease. *Lancet* 1977;2:523–526.
23. Rooney C, Sethi T. The epithelial cell and lung cancer: the link between chronic obstructive pulmonary disease and lung cancer. *Respiration* 2011;81:89–104.
24. Schottenfeld D, Beebe-Dimmer J. Chronic inflammation: a common and important factor in the pathogenesis of neoplasia. *CA Cancer J Clin* 2006;56:69–83.
25. Schwartz AG, Ruckdeschel JC. Familial lung cancer: genetic susceptibility and relationship to chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2006;173:16–22.
26. Engels EA. Inflammation in the development of lung cancer: epidemiological evidence. *Expert Rev Anticancer Ther* 2008;8:605–615.
27. Lee G, Walser TC, Dubinett SM. Chronic inflammation, chronic obstructive pulmonary disease, and lung cancer. *Curr Opin Pulm Med* 2009;15:303–307.
28. Saetta M, Di Stefano A, Turato G, et al. CD8+ T-lymphocytes in peripheral airways of smokers with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;157(3 Pt 1):822–826.
29. Barta P, Van Pelt C, Men T, Dickey BF, Lotan R, Moghaddam SJ. Enhancement of lung tumorigenesis in a Gprc5a Knockout mouse by chronic extrinsic airway inflammation. *Mol Cancer* 2012;11:4.
30. Hallström T, Singh B, Resman F, Blom AM, Mörgelin M, Riesbeck K. Haemophilus influenzae protein E binds to the extracellular matrix by concurrently interacting with laminin and vitronectin. *J Infect Dis* 2011;204:1065–1074.
31. Cheng Y, Lotan R. Molecular cloning and characterization of a novel retinoic acid-inducible gene that encodes a putative G protein-coupled receptor. *J Biol Chem* 1998;273:35008–35015.
32. Ye X, Tao Q, Wang Y, Cheng Y, Lotan R. Mechanisms underlying the induction of the putative human tumor suppressor GPRC5A by retinoic acid. *Cancer Biol Ther* 2009;8:951–962.
33. Xu XC, Sozzi G, Lee JS, et al. Suppression of retinoic acid receptor beta in non-small-cell lung cancer in vivo: implications for lung cancer development. *J Natl Cancer Inst* 1997;89:624–629.
34. Xu XC, Lee JS, Lee JJ, et al. Nuclear retinoid acid receptor beta in bronchial epithelium of smokers before and during chemoprevention. *J Natl Cancer Inst* 1999;91:1317–1321.
35. Hind M, Gilthorpe A, Stinchcombe S, Maden M. Retinoid induction of alveolar regeneration: from mice to man? *Thorax* 2009;64:451–457.

Published OnlineFirst October 19, 2012; DOI: 10.1158/1940-6207.CAPR-12-0290

Cancer Prevention Research

Research Article

See related commentary by Gomperts et al., p. 4

Characterizing the Molecular Spatial and Temporal Field of Injury in Early-Stage Smoker Non–Small Cell Lung Cancer Patients after Definitive Surgery by Expression Profiling

Humam Kadara[1], Li Shen[2], Junya Fujimoto[1], Pierre Saintigny[1], Chi-Wan Chow[1], Wenhua Lang[1], Zuoming Chu[1], Melinda Garcia[1], Mohamed Kabbout[1], You-Hong Fan[1], Carmen Behrens[1], Diane A. Liu[4], Li Mao[5], J. Jack Lee[4], Kathryn A. Gold[1], Jing Wang[2], Kevin R. Coombes[2], Edward S. Kim[1], Waun Ki Hong[1], and Ignacio I. Wistuba[1,3]

Abstract

Gene expression alterations in response to cigarette smoke have been characterized in normal-appearing bronchial epithelium of healthy smokers, and it has been suggested that adjacent histologically normal tissue displays tumor-associated molecular abnormalities. We sought to delineate the spatial and temporal molecular lung field of injury in smoker patients with early-stage non–small cell lung cancer (NSCLC; $n = 19$) who were accrued into a surveillance clinical trial for annual follow-up and bronchoscopies within 1 year after definitive surgery. Bronchial brushings and biopsies were obtained from six different sites in the lung at the time of inclusion in the study and at 12, 24, and 36 months after the first time point. Affymetrix Human Gene 1.0 ST arrays were used for whole-transcript expression profiling of airways ($n = 391$). Microarray analysis identified gene features ($n = 1,165$) that were nonuniform by site and differentially expressed between airways adjacent to tumors relative to more distant samples as well as those ($n = 1,395$) that were significantly altered with time up to 3 years. In addition, gene interaction networks mediated by phosphoinositide 3-kinase (PI3K) and extracellular signal–regulated kinase (ERK)1/2 were modulated in adjacent compared with contralateral airways and the latter network with time. Furthermore, phosphorylated AKT and ERK1/2 immunohistochemical expression were significantly increased with time (nuclear pAKT, $P = 0.03$; cytoplasmic pAKT, $P < 0.0001$; pERK1/2, $P = 0.02$) and elevated in adjacent compared with more distant airways (nuclear pAKT, $P = 0.04$; pERK1/2, $P = 0.03$). This study highlights spatial and temporal cancer–associated expression alterations in the molecular field of injury of patients with early-stage NSCLCs after definitive surgery that warrant further validation in independent studies. *Cancer Prev Res; 6(1); 8–17. ©2012 AACR.*

Introduction

Lung cancer, of which non–small cell lung cancer (NSCLC) comprises the majority, is the leading cause of cancer-related deaths in the United States and worldwide (1, 2). The high mortality of this disease is, in part, due to the late diagnosis of the majority of lung cancers after regional or distant spread of the malignancy (3). Recent data from the National Lung Screening Trial (4), indicating that screening increases early detection rates, are expected to augment the number of early-stage NSCLCs detected warranting the need for better clinical management of this growing subpopulation. Besides adjuvant therapy, there are no effective chemoprevention strategies for patients with early-stage NSCLCs (5) who comprise approximately 50% of all diagnosed cases and have a relatively high rate of relapse (3). Improved clinical management of early-stage NSCLCs is tightly linked to the identification of new effective early biomarkers that can guide potential chemoprevention strategies.

Most diagnosed NSCLC cases (85%) are attributable to cigarette smoking (6–8). Auerbach and colleagues found that cigarette smoke induces extensive histologic changes in the bronchial epithelia in the lungs of smokers and that premalignant lesions are widespread and multifocal throughout the respiratory epithelium, suggestive of a field effect (9). Many molecular abnormalities, such as LOH

Authors' Affiliations: Departments of [1]Thoracic/Head and Neck Medical Oncology, [2]Bioinformatics, [3]Pathology, and [4]Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas; and [5]School of Dentistry, The University of Maryland, Baltimore, Baltimore, Maryland

Note: Supplementary data for this article are available at Cancer Prevention Research Online (http://cancerprevres.aacrjournals.org/).

Current address for E.S. Kim: Department of Solid Tumor Oncology and Investigational Therapeutics, Levine Cancer Institute, Carolinas Healthcare System, Charlotte, North Carolina.

Corresponding Author: Ignacio I. Wistuba, Departments of Thoracic/Head and Neck Medical Oncology and Pathology, the University of Texas MD Anderson Cancer Center, Houston, TX 77030. Phone: 713-563-9184; Fax: 713-730-0309; E-mail: iiwistuba@mdanderson.org

doi: 10.1158/1940-6207.CAPR-12-0290

Published OnlineFirst October 19, 2012; DOI: 10.1158/1940-6207.CAPR-12-0290

(10–12), mutations in *TP53* (13), methylation of *p16* tumor suppressor, death-associated protein kinase (*DAPK*), and retinoic acid receptor 2 beta (*RAR-β2*), were detected in bronchial epithelia of cancer-free former smokers (14–18), some of which persist for many years after smoking cessation (15). More recently, global mRNA and miRNA expression profiles have been described in the normal-appearing bronchial epithelium of healthy smokers (19, 20) that are different from those in nonsmokers. Moreover, expression and pathway signatures have been derived from normal bronchial epithelium of smokers that exhibited diagnostic properties (21, 22). Molecular changes involving LOH of chromosomal regions 3p (*DDUT* and *FHIT* genes), 9p (*CDKN2A*), genomic instability (increased microsatellite repeats), and *p16* methylation have been shown in histologically normal epithelium in patients with squamous cell carcinoma and in the sequence of pathogenesis of the disease (14, 23, 24). Moreover, Nelson and colleagues showed that *KRAS* is also mutated in histologically normal lung tissue adjacent to lung tumors (25). Furthermore, Tang and colleagues found higher rates of *EGFR* mutations in adjacent normal bronchial epithelia (NBE; refs. 26, 27) suggestive of a potential localized field effect.

It is plausible to assume that understanding early molecular aberrations in histologically normal smoke-damaged airway epithelium of early-stage patients would serve as a critical first step toward identification of biomarkers that can guide lung cancer prevention strategies. However, the global molecular airway field of injury in patients with early-stage NSCLCs, in particular after definitive surgery, is unknown. In this study, we used transcript-level expression profiling coupled with gene interaction network analysis and immunohistochemical (IHC) analysis to characterize, in-depth, site- and time-dependent global molecular alterations in airways of smoker patients with early-stage NSCLCs.

Materials and Methods

Patient population and airway epithelial cell collection

Early-stage (I/II), current or former smoker patients with NSCLCs with at least a 10-pack-year smoking history and without evidence of disease after definitive surgery were recruited into the Vanguard phase II surveillance clinical trial (clinical trial number NCT00352391) within 1 year from time of surgery. Patients were accrued between 2004 and 2008 and underwent frequent testing including chest X-rays, computed tomographic (CT) scans, laboratory work, serologies, flexible bronchoscopies, and airway biopsy collections within 1 year from surgery (average, 6 months; range, 1–12; first time point), and at months 12, 24, and 36 following the first time point. Bronchoscopies were conducted using white light or both white light and autofluorescence modalities. Biopsies were obtained from all potential anatomic locations and time points per patient. In total, there were 324 evaluable airway biopsies. Histologic assessment was conducted to determine whether malignant changes will occur during the time period. Once patients have completed 3 years of testing, they were followed until the study is completed. Patients were composed of former ($n = 16$) and current ($n = 3$) smokers. One of the 3 current smoker patients quit smoking 6 months before the 24-month time point. The clinicopathologic variables of patients in this study are summarized in Table 1. The study was approved by the Institutional Review Boards, and all participants provided written informed consents.

Bronchial airway epithelial cells were obtained from 5 to 6 different sites (main carina, 4 airways from 4 lobes, and the bronchial stump or main stem bronchus adjacent to the originally resected tumor and lobe; Fig. 1) at each time point using an Olympus fiberoptic bronchoscope (Olympus America Inc.) and cytobrushes (Cellebrity Endoscopic Cytobrush, Boston Scientific). Patients ($n = 19$) with samples/specimens available for analysis that were obtained serially up to either 24 or 36 months and from at least 5 different sites at each time point ($n = 391$ airway samples) were selected for the study. Epithelial cell content was confirmed by cytokeratin staining which yielded a 90% epithelial cell content. Brushes were immediately placed in serum-free RPMI medium on ice, vortexed gently to disperse epithelia into the media, and then removed. Samples were then immediately centrifuged, and cell pellets were resuspended in 1 mL of PBS. About 500 μL of the sample was then again centrifuged, and the pellet resuspended in 500 μL β-mercaptoethanol containing RLT buffer (Qiagen), homogenized, and stored in −80°C until further processing. Total RNA was isolated using the RNeasy Mini Kit according to the manufacturer's instructions (Qiagen).

RNA processing for microarrays

Total RNA samples were preprocessed for subsequent hybridization to expression arrays using the WT-Ovation and Encore Biotin Module from NuGEN Technologies Inc. (San Carlos, CA) according to the manufacturer's instructions. Briefly, the WT-Ovation Pico RNA amplification system was used to generate amplified cDNA using 5 ng of starting RNA material. After formation of double stranded cDNA, DNA was amplified using the SPIA Amplification Method, a linear isothermal DNA amplification process developed by the vendor (NuGEN Technologies). The WT-Ovation Exon Module (NuGEN Technologies) was then used for generation of amplified sense strand cDNA (ST-cDNA) that is suitable for subsequent array analysis with the Affymetrix Human Gene 1.0 ST array platform (Affymetrix, Santa Clara, CA). Fragmented and biotin-labeled cDNA was then generated using the Encore Biotin Module (NuGEN Technologies) using 5 μg of amplified cDNA. Quality and size distribution of unfragmented SPIA-amplified cDNA and subsequent fragmented labeled cDNA were assessed by loading samples on an RNA 6000 Nano LabChip (Agilent) and analysis with Agilent Bioanalyzer 2000 (Agilent). No differences in quality were noted on the basis of the duration of sample storage.

Generation of microarray raw data and analysis

Fragmented and labeled cDNA (2.5 μg) were hybridized onto Human Gene 1.0 ST arrays according to the

Published OnlineFirst October 19, 2012; DOI: 10.1158/1940-6207.CAPR-12-0290

Table 1. Clinicopathologic features of patients with NSCLCs included in the study

Patient	Histology	Anatomic site[a]	Stage	Recurrence	Adjuvant chemotherapy	Gender/ age, y	Smoking status[b]	Pack-years	Years quit smoking[c]	Months to inclusion from surgery[d]
1	ADC	LUL	IIA	Yes	No	Male/81	Former	100	12.0	1
2	ADC	RLL	IB	No	No	Male/64	Former	27	19.1	5
3	ADC	RUL	IB	Yes	Yes	Male/58	Former	12	25.5	7
4	SCC	RLL	IB	Yes	Yes	Female/62	Former	60	7.6	7
5	ADC	RLL	IA	No	No	Female/60	Former	31	11.8	6
6	ADC	LLL	IA	Yes	Yes	Male/62	Former	40	5.8	5
7	ADC	RUL	IB	Yes	Yes	Male/61	Current	70	NA	10
8	SCC	LUL	IB	No	No	Male/64	Current	100	NA	5
9	ADC	RUL	IA	No	No	Male/69	Former	12	36.5	8
10	ADC	RUL	IB	No	No	Female/63	Former	18	8.6	5
11	ADC	RUL	IB	Yes	No	Female/53	Former	42	5.8	3
12	SCC	RUL	IA	No	No	Male/62	Former	68	0.4	5
13	SCC	LUL	IA	No	No	Male/70	Former	96	1.4	7
14	ADC	LUL	IA	Yes	Yes	Female/45	Former	12.5	0.9	12
15	ADC	RLL	IA	No	No	Female/57	Current	80	NA	4
16	ADC	LLL	IA	No	No	Male/65	Former	48	21.0	7
17	ADC	LLL	IA	No	No	Male/64	Former	92	5.1	7
18	ADC	RML	IB	No	No	Male/71	Former	102	2.3	2
19	ADC	LUL	IA	No	No	Female/66	Former	50	18.5	5

Abbreviations: ADC, adenocarcinoma; LLL, left lower lobe; LUL, left upper lobe; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe; SCC, squamous cell carcinoma.

[a]Location of primary tumor in the lung.

[b]Smoking status at time of inclusion into the study (patient 7 quit smoking during the course of the study).

[c]Years from smoking cessation to time of inclusion into the study.

[d]Months elapsed from surgery to time of inclusion into the study/first bronchoscopy time point.

manufacturer's instructions (Affymetrix). Hybridization cocktails containing samples, control oligonucleotide and eukaryotic hybridization controls in addition to hybridization mixes, dimethyl sulfoxide and nuclease-free water were heat denatured at 99°C for 5 minutes, cooled to 45°C for 5 minutes, and finally centrifuged at maximum speed for 1 minute. After injecting 80 μL of the hybridization cocktails, arrays were incubated for 17 hours in a hybridization oven set to a temperature of 45°C with 60 rpm rotation. Arrays were washed, stained, and processed using Affymetrix GeneChip Fluidics Station 450 systems after which they were imaged using Affymetrix GeneChip Scanner 3000 7G for subsequent generation of raw data (*CEL files). Raw data were quantified using Robust Multichip Array (RMA) background correction, quantile normalization, and RMA probe-level models (RmaPlm) and summarization methods. MIAME compliant metadata, normalized expression values, and 391 CEL files were submitted to the Gene Expression Omnibus (GEO; samples GSM992943-GSM993345, series GSE40407). After data preprocessing and normalization, a $\log_2$ transformation was applied. Pathways and gene interaction network analyses were conducted using the commercially available software Ingenuity Pathways Analysis (IPA, Redwood City, CA). All details of the microarray analysis including pairwise analysis of adjacent and contralateral airways for patient clustering are included in the Supplementary Information and in the 4 Supplementary Sweave Reports accompanying the manuscript.

IHC analysis of airway biopsies

Immunohistochemistry was done on histologic sections of 4-μm formalin-fixed, paraffin-embedded tissue samples prepared by a tissue arrayer as described previously (28). IHC analysis was conducted using purified rabbit polyclonal primary antibodies raised against phospho-AKT(threonine308) (1:200 dilution, clone C31E5E, catalog number 2965) and phospho-ERK1/2(Thr202/Tyr204) (1:400 dilution, clone D13.14.4E, catalog number 4370; Cell Signaling Technology). Antigen retrieval was conducted using the Dako Target Retrieval System at a pH of 6 (Dako). Intrinsic peroxidase activity was blocked by 3% methanol and hydrogen peroxide for 15 minutes and serum-free protein block (Dako) was used for 7 minutes for blocking nonspecific antibody binding. Slides were then incubated with the antibodies against phospho-AKT and phospho-ERK1/2 at room temperature for 90 and 120 minutes, respectively. After 3 washes in TBS, slides were incubated for 30 minutes with Dako Envision+ Dual Link at room temperature.

Published OnlineFirst October 19, 2012; DOI: 10.1158/1940-6207.CAPR-12-0290

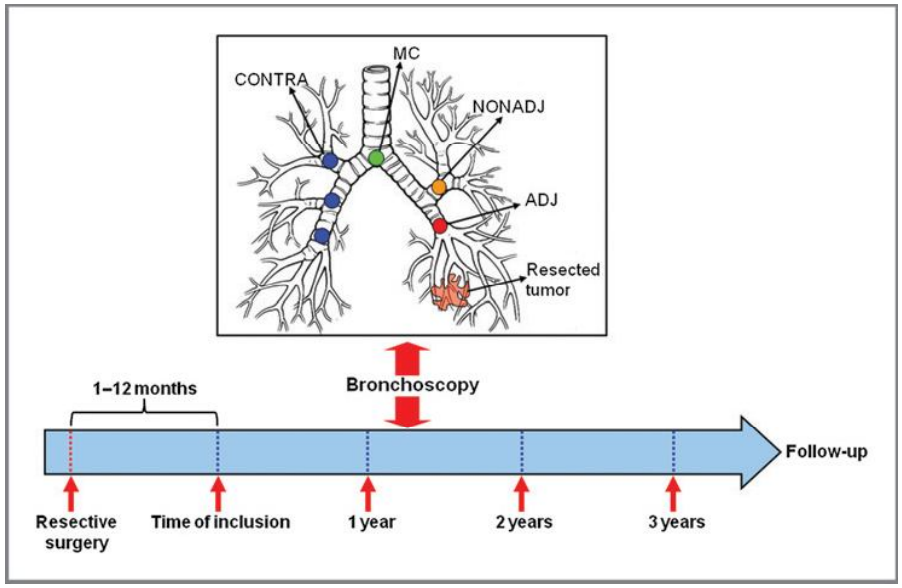


Figure 1. Spatial and temporal molecular field of injury in patients with early-stage NSCLCs after definitive surgery. Schematic depicting the site-dependent (top) and time-dependent (bottom) collection of airway epithelia brushings by bronchoscopy. Smoker patients with early-stage NSCLCs were enrolled into a surveillance clinical trial for annual follow-up and bronchoscopies within 1 year after definitive surgery. Bronchial airway epithelial cells (brushings and biopsies) were obtained from 5 to 6 different sites composed of the main carina (MC), 4 airways from 4 lobes ipsilateral (NON-ADJ), or contralateral (CONTRA) to the originally resected tumor and of the bronchial stump or main stem bronchus adjacent to the tumor and lobe. All site-different airway epithelia were collected at the time of inclusion in the study and at 12, 24, and 36 months following the starting time point (391 airways from 19 patients with NSCLCs).

Following 3 additional washes, slides were incubated with Dako chromogen substrate for 5 minutes and were counterstained with hematoxylin for another 5 minutes. Formalin-fixed, paraffin-embedded pellets from lung cancer cell lines displaying positive phospho-AKT and phospho-ERK1/2 expression by Western blot analysis were used as a positive control, whereas samples and whole-section tissue specimens processed similarly, except for the omission of the primary antibodies used as negative controls. The intensity and extent of cytoplasmic and nuclear phospho-AKT and phospho-ERK1/2 immunostaining were evaluated using a light microscope (magnification, $\times 20$) independently by 2 pathologists (J. Fujimoto and I.I. Wistuba) who were blinded to the identity of the samples. Cytoplasmic expression was quantified using a 4-value intensity score (0, none; 1, weak; 2, moderate; and 3, strong) and the percentage (0%–100%) of the extent of reactivity. A final cytoplasmic expression score was obtained by multiplying the intensity and reactivity extension values (range, 0–300). Nuclear expression score was quantified using the percentage of extent of reactivity (range, 0–100).

Summary statistics, including frequency tabulation, means, SDs, median, and range, were given to describe subject characteristics and IHC protein expression. Repeated measures analysis was conducted to assess the differential effect of sites on phosphorylated AKT and extracellular signal–regulated kinase (ERK) 1/2 expression variation with time. Mixed-effects models were generated to assess significance of site, time, and the interaction of both factors to the expression of both proteins. All statistical tests were 2-sided and $P \leq 0.05$ was considered to be statistically significant. Statistical analysis was conducted with standard statistical software, including SAS Release 8.1 (SAS Institute) and S-Plus 2000 (Mathsoft Inc.).

Results

Detailed site- and time-dependent airway sampling of the field of injury in early-stage NSCLC patients after definitive surgery

Expression patterns molecularly exemplifying the impact of smoking on the airway epithelium of cancer-free individuals have been characterized (8, 19, 29). Moreover, molecular abnormalities typically found in lung tumors have been detected in normal resected margins suggestive of a field effect (15, 25–27, 30, 31). However, the biologic nature of the field of injury in particular after complete removal of the tumor in patients with early-stage NSCLCs, who are increasing in number and for whom there are no chemoprevention strategies, are yet unknown. Smoker patients with early-stage NSCLCs were recruited on a prospective phase II surveillance clinical trial that included frequent CT scans, serologies, and annual bronchoscopies in which airway brushings and biopsies were obtained within 1 year following tumor resection and at 12, 24, and 36 months following the first time point (Fig. 1). The first time point bronchoscopies were all conducted within 1 year of definitive surgery (average, 6 months; range, 1–12). Nineteen patients were selected for the study (Table 1) on the basis of airway sampling of at least 5 different sites per time point

Published OnlineFirst October 19, 2012; DOI: 10.1158/1940-6207.CAPR-12-0290

and continuously up to 24 or 36 months giving rise to 391 airway samples for transcript-level expressing profiling. The patients were accrued between 2004 and 2008 and were composed of former ($n = 16$) and current ($n = 3$) smokers. Brushings and biopsies were obtained from the main carina, airways relatively adjacent to the previously resected tumor and from airways more distant from the tumor in the ipsilateral and contralateral lung (Fig. 1).

Following normalization of the raw expression data, a mixed-effects model was applied to characterize the expression pattern of each gene that incorporated fixed-effects such as the site from the tumor and time after surgery of the collected airway samples (Supplementary Information). Histogram *P* value distribution plots after fitting beta-uniform mixture (BUM) models (32) for derived *P* values on the basis of the fixed-effects suggested that both site and time of the airway samples influenced gene expression modulation (Supplementary Fig. S1A and S1B).

Site-dependent differential expression patterns in airway epithelia of early-stage NSCLC patients

We first sought to determine whether gene expression profiles are differentially expressed in airways by site from the tumor including those relatively adjacent to the resected tumors compared with more distant airways. On the basis of the generated BUM models and *P* value distributions, genes differentially expressed by site were selected on the basis of a 1% false discovery rate (FDR). We identified 1,165 gene features that were statistically significantly differentially expressed by site (Supplementary Table S1 and Supplementary Sweave Report S2). Two-dimensional hierarchical clustering showed that the airway samples were divided into 2 main branches or clusters (Fig. 2A) on the basis of expression of the genes. Moreover, the left cluster in the indicated heatmap's dendogram contained a statistically significantly higher number of adjacent airway samples than the right branch which contained a significantly higher proportion of main carinas and contralateral to the tumor airways ($P = 0.0027$ of the Fisher exact test for count data). In addition, the 2-dimensional clustering revealed 8 different gene expression patterns which are indicated by the color bar and code along the left side of the heatmap (Fig. 2A). Some of the different gene clusters or classes were associated with a specific group of airway samples. Notably, a cluster of 263 genes (cluster 1, Supplementary Table S1), indicated by the dark green color and asterisk on the heatmap, was found to have highest average expression in adjacent airways (Fig. 2A and Supplementary Table S2). In contrast, another cluster of 348 genes (indicated by magenta color) appeared to be highest in expression in main carinas (Fig. 2A).

We then determined to functionally analyze the cluster of genes ($n = 263$) that was found to exhibit the highest average expression in adjacent airways between adjacent and contralateral airways. Functional pathways analysis using IPA depicted several significantly modulated

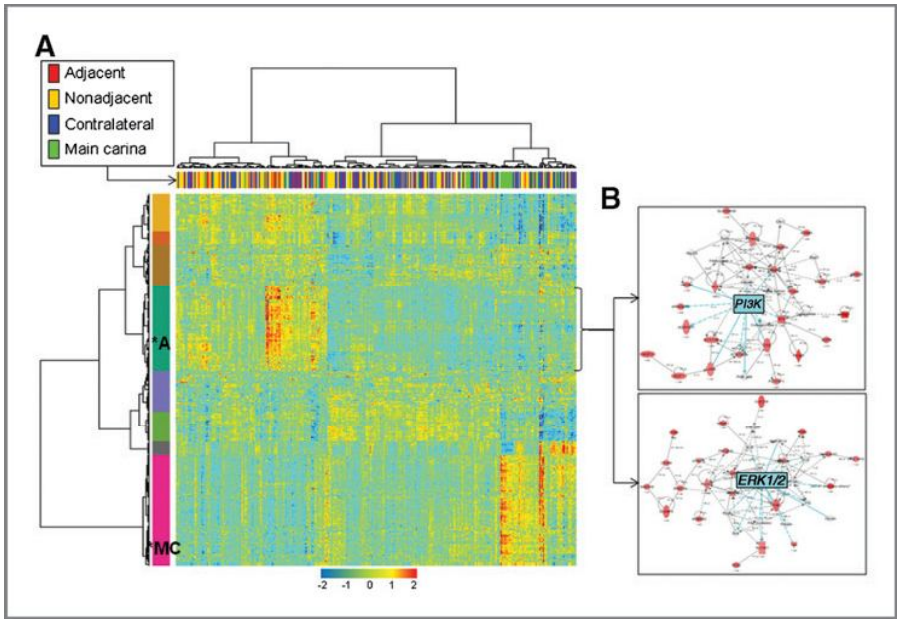


Figure 2. Site-dependent airway epithelia differential gene expression patterns. A, heatmap depicting 2-dimensional clustering of airway samples ($n = 391$) and genes ($n = 1,165$) that were determined to be differentially expressed by site in the mixed-effects model on the basis of a 1% FDR cutoff. The identified 8 gene clusters are labeled with different colors with the green cluster of genes ($n = 263$, *A) exhibiting highest average expression in adjacent airways and the magenta cluster ($n = 348$, *MC) having highest expression in main carinas. B, gene interaction analysis by IPA depicting networks with significant scores that indicate the likelihood of genes in a network being found together than due to chance. The depicted networks were found to be mediated by *PI3K* (top) and *ERK1/2* (bottom) with both kinases themselves not modulated in expression. Gene expression variation based on the statistical cutoff described above is depicted by color in the network (red, upregulated; green, downregulated).

Published OnlineFirst October 19, 2012; DOI: 10.1158/1940-6207.CAPR-12-0290

pathways and molecular functions in particular, T cell ($P = 1.2 \times 10^{-9}$), chemokine C-C motif receptor 5 (CCR5; $P = 1.5 \times 10^{-9}$), and phospholipase C signaling ($P = 3.6 \times 10^{-9}$). In addition, topological gene interaction network analysis highlighted functionally modulated and upregulated gene networks mediated by phosphoinositide 3-kinase (*PI3K*) and *ERK* in the adjacent airways (Fig. 2B). It is worthwhile to note that we also observed increased modulation of *ERK*/mitogen-activated protein kinase (*MAPK*)-mediated gene interaction network using a different analytic method where we compared expression profiles between adjacent and contralateral airways between patients in a pairwise fashion (data not shown, Supplementary Sweave Report S3). These findings suggest that airway site-dependent differential gene expression profiles in patients with early-stage NSCLCs exhibit increased molecular features and gene interaction networks typically associated with cancers.

We then determined to analyze gene expression profiles while excluding main carinas because their epithelial anatomy is suggested to be different from that of other airways (33) and thus may confound site-dependent observations. Following exclusion of main carinas, we found a reduced number of genes ($n = 136$) that were significantly modulated by site in the field of injury (Supplementary Sweave Report S4). Two-dimensional hierarchical clustering showed that airway samples were divided into 2 main branches or clusters (Supplementary Fig. S2) with significantly more adjacent airways in the right branch ($P = 0.0002$ of the Fisher exact t test). Moreover, the differentially expressed genes (Supplementary Table S4) were composed of 2 main subgroups with one cluster (top cluster) of 113 genes exhibiting highest average expression in adjacent airways. It is important to note that when we cross-compared gene clusters that we had found to exhibit the highest average expression in adjacent airways when including (263-gene cluster) or excluding (113-gene cluster) main carinas, we found a highly significant overlap ($P = 2.46 \times 10^{-191}$) in the number of genes ($n = 96$). Moreover, the site-dependent genes identified after exclusion of main carinas were also topologically organized following functional pathways analysis into interaction networks involving PI3K and ERK.

Gene expression profiles in the lung airway epithelia of early-stage NSCLC patients are modulated with time following definitive surgery

We then determined to identify genes that were differentially expressed with time. On the basis of the generated BUM models and P value distributions, time-dependent differentially expressed genes were identified and selected on the basis of a 5% FDR cutoff ($n = 1,395$; Supplementary Table S3 and Supplementary Sweave Report S2). Hierarchical clustering of samples indicated that the dendogram's main branches were statistically significantly unbalanced with respect to time; the main left branch comprised a higher number of 24- and 36-month time points than the right cluster ($P = 4.15 \times 10^{-7}$ of the Fisher exact test for count data). Two-dimensional clustering of both genes and samples revealed 2 main classes of genes: those that displayed increased (upper cluster) and decreased (lower cluster) expression with time (Fig. 3A). Functional pathways and gene interaction network analysis of genes differentially expressed between 36 months and the first time point revealed statistically significantly modulated pathways, in particular protein ubiquitination (5.3×10^{-5}), glutathione metabolism (6.0×10^{-5}), mitochondrial dysfunction (1.4×10^{-4}), and oxidative phosphorylation (2.9×10^{-3}) as well as eukaryotic initiation factor 2 (eIF2) signaling (2.6×10^{-3}). In addition, network analysis highlighted functionally modulated and elevated gene interaction

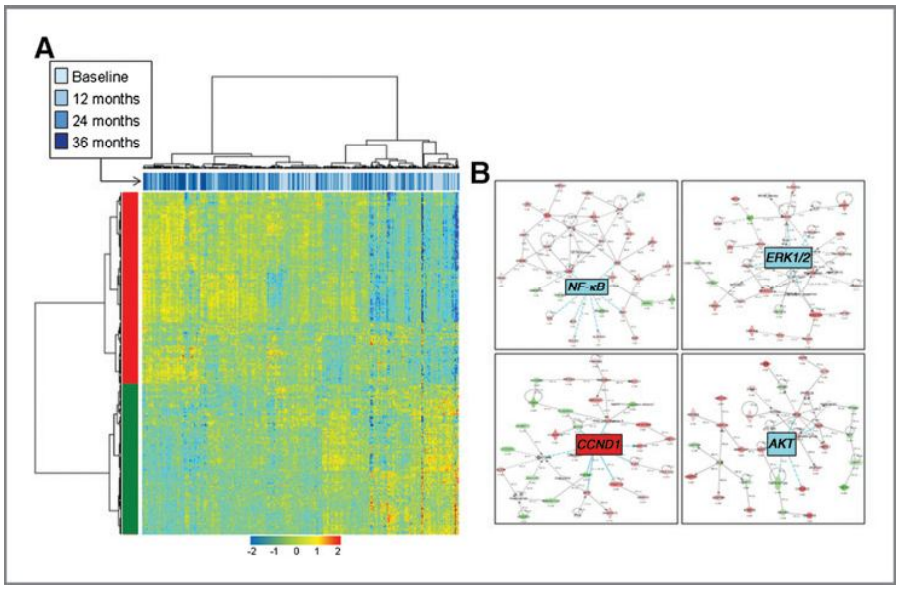


Figure 3. Temporal modulation of the molecular field of injury after definitive surgery in patients with early-stage NSCLCs. A, heatmap depicting 2-dimensional clustering of airway samples ($n = 391$) and genes ($n = 1,395$) that were determined to be differentially expressed by time in the mixed-effects model on the basis of a 5% FDR cutoff. B, gene interaction analysis, similar to that in Fig. 2, by IPA depicting networks with increased likelihood of genes being found together than due to chance and mediated by *NF-κB*, *ERK1/2*, *AKT*, and *CCND1*. *CCND1* itself was upregulated at the expression level. Gene expression variation based on the statistical cutoff is depicted by color in the network (red, upregulated; green, downregulated).

Published OnlineFirst October 19, 2012; DOI: 10.1158/1940-6207.CAPR-12-0290

networks with time in particular those mediated by *NF-κB*, *ERK*, *AKT*, and cyclin-B1 (*CCNB1*; Fig. 3B).

We then sought to determine the relationship of genes that were significantly modulated by site and time in the molecular field of injury. A smooth scatter plot of transformed *P* values indicated that site- and time-dependent expression modulations were largely independent (Supplementary Fig. S3 and Supplementary Sweave Report S2). We then cross-compared the site-dependent ($n = 1{,}165$) and time-dependent ($n = 1{,}395$) profiles that we had noted in the molecular field of injury. Using hypergeometric tests for overlapping probability, we found no significant overlap between genes we had determined to be differentially expressed by site and time in the molecular field of injury ($P = 0.865$; Supplementary Sweave Report S4).

Increased expression of phosphorylated AKT and ERK1/2 in airway epithelia by site from the tumor and with time following surgery

Our findings on the modulation of PI3K- and ERK1/2-mediated networks by site and time after surgery prompted us to examine the IHC expression of surrogate markers of both signaling cascades in corresponding formalin-fixed, paraffin-embedded airway biopsies. We sought to examine expression of phosphorylated AKT at Threonine308 because phosphorylation of this amino acid is well known to occur through phosphoinositide-dependent kinase 1 (PDPK1) following PI3K activation (34). We assessed by immunohistochemistry the IHC expression of phospho-AKT (Thr308) and phospho-ERK1/2(Thr202/Tyr204) in available and evaluable histologically NBE biopsies ($n = 324$) corresponding to the brushings analyzed by expression profiling. Immunoreactivity of phospho-AKT (minimum, 0; maximum, 300) was variable as depicted in the representative photomicrographs in Fig. 4A and was detected in both the cytoplasm and nucleus of NBE (Fig. 4A). IHC analysis showed that cytoplasmic ($P < 0.0001$) and nuclear ($P = 0.01$) phospho-AKT statistically significantly increased with time up to 3 years in all NBE (Fig. 4B) with highest expression at the 36-month time point. Nuclear phosphorylated AKT was also statistically significantly increased in adjacent NBE compared with airways more distant from tumors in the mixed-effects model ($P = 0.04$; Fig. 4B).

Immunoreactivity of phospho-ERK1/2 was also variable (minimum, 0; maximum, 209) and mainly cytoplasmic (Fig. 4C). IHC analysis showed that phosphorylated ERK1/2 was statistically significantly elevated in adjacent NBE ($P = 0.03$) and significantly increased with time up to 3 years in all airways when averaged together ($P = 0.02$; Fig. 4D) in the model. Notably, there was a significant interaction ($P = 0.019$) between the site of NBE and the time of sampling, as phospho-ERK1/2 expression was significantly increased with time in adjacent NBE but not in contralateral airways and main carinas in the model (Fig. 4D) with highest

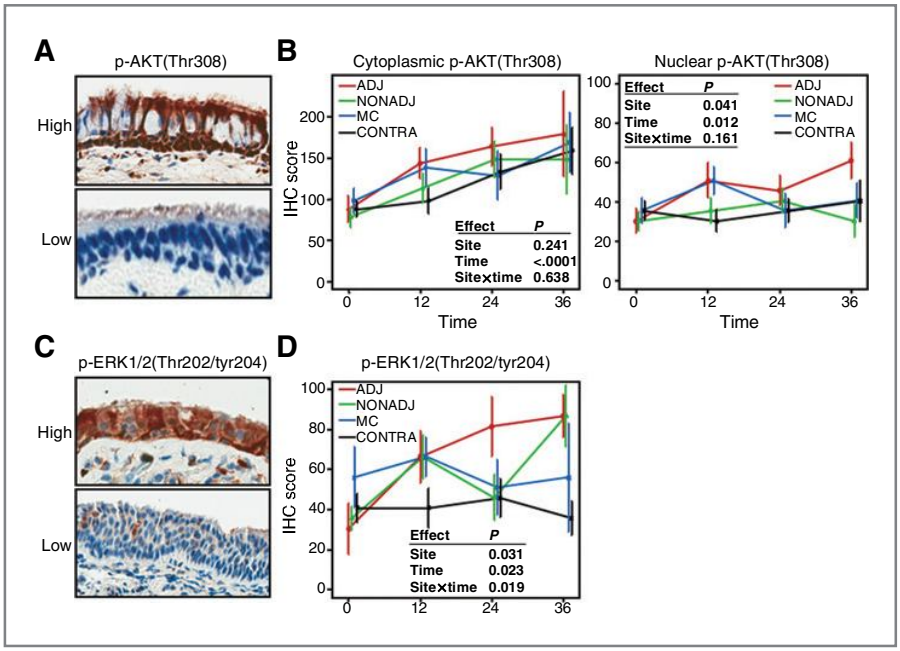


Figure 4. Site- and time-dependent IHC expression of phosphorylated AKT and ERK1/2 in the airway field of injury. A, representative photomicrographs (magnification, ×20) depicting strong (top) and weak (bottom) phospho-AKT(Thr308) immunostaining. B, IHC scores of cytoplasmic (left) and nuclear phospho-AKT (middle) were assessed for statistically significant differences by site and time in a mixed-effect model and plotted in main carinas (MC) and in adjacent (ADJ), nonadjacent (ipsilateral, NON-ADJ), and contralateral (CONTRA) airways with time. C, representative photomicrographs (magnification, ×20) depicting strong (top) and weak (bottom) phospho-ERK1/2(Thr202/Tyr204) immunostaining. D, IHC scores of phosphorylated ERK1/2 levels were assessed for statistically significant differences by site and time in a mixed-effect model and repeated measures analysis, site × time, term for interaction between site and time factors. Error bars represent SEM.

Published OnlineFirst October 19, 2012; DOI: 10.1158/1940-6207.CAPR-12-0290

expression in adjacent airways and nonadjacent (ipsilateral, green plot) airways observed at the latest time point. Similar data were obtained when we excluded main carinas in the mixed-effects model (data not shown). Moreover, we noted similar findings when all airway samples were analyzed irrespective of the presence of preneoplastic lesions (e.g., dysplasias).

These data showed that like differential gene expression profiles within the lung airway field of injury, canonical activated oncogenes are modulated by site from the resected tumor and time following definitive surgery in patients with early-stage NSCLCs.

Discussion

In this report, we characterized differential expression profiles and protein expression within the lung airway field of injury of patients with early-stage NSCLCs by site from the tumor and time in years following surgery. We showed, and to our knowledge for the first time, that gene expression profiles in histologically normal airways of patients with early-stage NSCLCs are nonuniform by site and are modulated with time up to 3 years following surgery. Moreover, functional analysis of the expression profiles showed that canonical expression patterns and protein kinase activation, typical of tumors, are increased in airway sites adjacent to tumors as well as remain or are further modulated in the lung airway field of injury for three years after definitive surgery. In particular, phosphorylated ERK/MAPK and AKT expression were upregulated in NBE with time and by site from tumors. In light of the growing subpopulation of early-stage NSCLCs, our findings are, in part, proof-of-principle and raise the intriguing possibility of the importance of intense surveillance and molecular characterization of the remaining smoking-injured airway epithelia and its potential integration in the future into clinical practice and management of early-stage disease. However, it is noteworthy that our study's patient cohort, despite its uniqueness in which expression profiling was serially conducted on airways from multiple sites collected during bronchoscopies for 36 months following surgery, is of limited size. Moreover, the reported findings warrant the need for validation or confirmation in independent larger sets.

There is a large body of evidence that patients who have survived an upper aerodigestive cancer comprise a high-risk population that may be targeted for early detection and chemoprevention efforts (5, 31). Currently, there are no established adjuvant treatments in the tertiary prevention setting of patients with early-stage NSCLCs. It has been suggested that failures in advances of chemoprevention are, in part, due to the lack of clear and specific molecular targets (5, 35). Our extensive profiling of the molecular field of injury in patients with NSCLCs identified cancer-associated pathways (PI3K and ERK) aberrantly regulated in NBE of patients with NSCLCs after tumor removal. In this context, it is plausible to suggest that a thorough characterization of the molecular field of injury in patients with early-stage NSCLCs can aid in identification of aberrantly expressed pathways, for example, PI3K and ERK, which could potentially serve as suitable targets for chemoprevention. However, it cannot be neglected that the alternative hypothesis can counter argue that activation of PI3K and ERK/MAPK pathways may be beneficial for chemoprevention as markers of both pathways (phosphorylated AKT and ERK) increased following surgical tumor resection. Our suggestion that such pathways may serve as chemoprevention targets should be interpreted cautiously and is presented owing to the known promalignant function of the pathways and gene networks highlighted in our analysis (36).

The first time point brushings in this study were obtained within 1 year of patient accrual and not at the time when the tumor was still *in situ*. The variable starting time point from surgery among patients is a limiting factor in our analysis as it is plausible to assume that the molecular field of injury may vary upon tumor removal. We were not able to avoid this caveat given the difficulty of accrual of patients following surgical tumor resection to obtain bronchial brushings at 6 different anatomical sites in the bronchial tree annually up to 3 years. However, it is important to mention that the time-dependent gene expression profiles we had identified, despite not incorporating the molecular field effect at the time the tumor was present, showed gradual changes in expression with time. This effect was also evident when analyzing the IHC expression of phosphorylated AKT and ERK in corresponding bronchial biopsies with highest expression at 3 years. Our findings warrant future larger studies in which the molecular field of injury at the time the tumor is still *in situ* can also be serially monitored for several years.

The patient population we had studied was composed of patients with early-stage NSCLCs, which is in contrast to earlier transcriptomic studies of the molecular field of injury that focused on phenotypically normal smokers and nonsmokers (19, 21, 22). It is still not clear whether the differences in expression described in this study reflect an already present gradient field of injury that may have contributed to tumor development in light of the differential cancer-associated pathways identified or one that arises due to the molecular impact of the tumor on the adjacent field. It is important to note that in this study, the spatial and temporal molecular field of injury in patients with lung cancer was profiled prospectively starting within 1 year following definitive surgery. Thus, the above speculation may be addressed by a similar thorough spatial and temporal characterization of the molecular field of injury before and after surgery in early-stage patients. In addition, we did not have access to similar type of brushings from cancer-free individuals such as high-risk smokers. Similar analysis of the molecular field of injury in cancer-free individuals will shed light on the nature of site- and time-dependent expression patterns in the field of injury and whether such changes in patients with cancer reflect recovery from surgery (temporal profiles) or are a cause or consequence of tumor development in the adjacent field (spatial profiles).

In conclusion, our unique study identified gene expression profiles, functional gene networks, and activated levels

Published OnlineFirst October 19, 2012; DOI: 10.1158/1940-6207.CAPR-12-0290

of oncogenic protein kinases within the field of injury of patients with early-stage NSCLCs that are modulated or increased in airways spatially from the tumor and temporally following surgery. Moreover, the herein previously uncharacterized airway cancer–associated expression and protein kinase alterations harbor potential valuable targets for chemoprevention and warrant confirmation and further studies in larger independent cohorts.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: H. Kadara, L. Mao, E.S. Kim, I.I. Wistuba
Development of methodology: H. Kadara, L. Shen, L. Mao, I.I. Wistuba
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): H. Kadara, J. Fujimoto, P. Saintigny, M. Kabbout, Y.-H. Fan, C. Behrens, K.A. Gold
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): H. Kadara, L. Shen, J. Fujimoto, M. Kabbout, D.A. Liu, J.J. Lee, J. Wang, K.R. Coombes, E.S. Kim, I.I. Wistuba
Writing, review, and/or revision of the manuscript: H. Kadara, J. Fujimoto, P. Saintigny, J.J. Lee, K.A. Gold, K.R. Coombes, W.K. Hong, I.I. Wistuba
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J. Fujimoto, C.-W. Chow, W. Lang, Z. Chu, E.S. Kim, I.I. Wistuba
Study supervision: H. Kadara, L. Mao, E.S. Kim, W.K. Hong, I.I. Wistuba

Grant Support

This study was funded in part by Department of Defense (DoD) grants W81XWH-04-1-0142 (to W.K. Hong and I.I. Wistuba) and W81XWH-10-1-1007 (to H. Kadara and I.I. Wistuba) and by Cancer Center Support Grant CA16672 (MD Anderson Cancer Center microarray core facility).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 3, 2012; revised September 27, 2012; accepted September 28, 2012; published OnlineFirst October 19, 2012.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69–90.
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin 2012;62:10–29.
3. Herbst RS, Heymach JV, Lippman SM. Lung Cancer. N Engl J Med 2008;359:1367–80.
4. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365:395–409.
5. Gold KA, Kim ES, Lee JJ, Wistuba II, Farhangfar CJ, Hong WK. The BATTLE to personalize lung cancer prevention through reverse migration. Cancer Prev Res 2011;4:962–72.
6. Goldstraw P, Ball D, Jett JR, Le Chevalier T, Lim E, Nicholson AG, et al. Non-small-cell lung cancer. Lancet 2011;378:1727–40.
7. Gazdar AF, Thun MJ. Lung cancer, smoke exposure, and sex. J Clin Oncol 2007;25:469–71.
8. Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. Cancer Prev Res 2008;1:396–403.
9. Auerbach O, Stout AP, Hammond EC, Garfinkel L. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. N Engl J Med 1961;265:253–67.
10. Mao L, Lee JS, Kurie JM, Fan YH, Lippman SM, Lee JJ, et al. Clonal genetic alterations in the lungs of current and former smokers. J Natl Cancer Inst 1997;89:857–62.
11. Powell CA, Klares S, O'Connor G, Brody JS. Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer. Clin Cancer Res 1999;5:2025–34.
12. Wistuba II, Lam S, Behrens C, Virmani AK, Fong KM, LeRiche J, et al. Molecular damage in the bronchial epithelium of current and former smokers. J Natl Cancer Inst 1997;89:1366–73.
13. Franklin WA, Gazdar AF, Haney J, Wistuba II, La Rosa FG, Kennedy T, et al. Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. J Clin Invest 1997;100:2133–7.
14. Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, et al. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. Proc Natl Acad Sci U S A 1998;95:11891–6.
15. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res 2002;62:2370–7.
16. Heller G, Zielinski CC, Zochbauer-Muller S. Lung cancer: from single-gene methylation to methylome profiling. Cancer Metastasis Rev 2010;29:95–107.
17. Soria JC, Rodriguez M, Liu DD, Lee JJ, Hong WK, Mao L. Aberrant promoter methylation of multiple genes in bronchial brush samples from former cigarette smokers. Cancer Res 2002;62:351–5.
18. Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, Seidl S, et al. Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. Int J Cancer 2003;107:612–6.
19. Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. Proc Natl Acad Sci U S A 2004;101:10143–8.
20. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, et al. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. Proc Natl Acad Sci U S A 2009;106:2319–24.
21. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nat Med 2007;13:361–6.
22. Gustafson AM, Soldi R, Anderlind C, Scholand MB, Qian J, Zhang X, et al. Airway PI3K pathway activation is an early and reversible event in lung cancer development. Sci Transl Med 2010;2:26ra5.
23. Wistuba II, Behrens C, Virmani AK, Mele G, Milchgrub S, Girard L, et al. High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. Cancer Res 2000;60:1949–60.
24. Wistuba II, Behrens C, Virmani AK, Milchgrub S, Syed S, Lam S, et al. Allelic losses at chromosome 8p21-23 are early and frequent events in the pathogenesis of lung cancer. Cancer Res 1999;59:1973–9.
25. Nelson MA, Wymer J, Clements N Jr. Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers. Cancer Lett 1996;103:115–21.
26. Tang X, Shigematsu H, Bekele BN, Roth JA, Minna JD, Hong WK, et al. EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. Cancer Res 2005;65:7568–72.
27. Tang X, Varella-Garcia M, Xavier AC, Massarelli E, Ozburn N, Moran C, et al. Epidermal growth factor receptor abnormalities in the pathogenesis and progression of lung adenocarcinomas. Cancer Prev Res 2008;1:192–200.
28. Sun M, Behrens C, Feng L, Ozburn N, Tang X, Yin G, et al. HER family receptor abnormalities in lung cancer brain metastases and corresponding primary tumors. Clin Cancer Res 2009;15:4829–37.
29. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. Genome Biol 2007;8:R201.

Published OnlineFirst October 19, 2012; DOI: 10.1158/1940-6207.CAPR-12-0290

30. Wistuba II, Behrens C, Milchgrub S, Bryant D, Hung J, Minna JD, et al. Sequential molecular abnormalities are involved in the multistage development of squamous cell lung carcinoma. Oncogene 1999;18:643–50.
31. Wistuba II, Gazdar AF. Lung cancer preneoplasia. Annu Rev Pathol 2006;1:331–48.
32. Pounds S, Morris SW. Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. Bioinformatics 2003;19:1236–42.
33. Lips KS, Volk C, Schmitt BM, Pfeil U, Arndt P, Miska D, et al. Polyspecific cation transporters mediate luminal release of acetylcholine from bronchial epithelium. Am J Respir Cell Mol Biol 2005;33:79–88.
34. Engelman JA. Targeting PI3K signalling in cancer: opportunities, challenges and limitations. Nat Rev Cancer 2009;9:550–62.
35. Khuri FR. The dawn of a revolution in personalized lung cancer prevention. Cancer Prev Res 2011;4:949–53.
36. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–74.

Cancer Prevention Research

Characterizing the Molecular Spatial and Temporal Field of Injury in Early-Stage Smoker Non –Small Cell Lung Cancer Patients after Definitive Surgery by Expression Profiling

Humam Kadara, Li Shen, Junya Fujimoto, et al.

Cancer Prev Res 2013;6:8-17. Published OnlineFirst October 19, 2012.

DOI:10.1093/jnci/dju004
First published online February 22, 2014

ARTICLE

Transcriptomic Architecture of the Adjacent Airway Field Cancerization in Non–Small Cell Lung Cancer

Humam Kadara, Junya Fujimoto, Suk-Young Yoo, Yuho Maki, Adam C. Gower, Mohamed Kabbout, Melinda M. Garcia, Chi-Wan Chow, Zuoming Chu, Gabriella Mendoza, Li Shen, Neda Kalhor, Waun Ki Hong, Cesar Moran, Jing Wang, Avrum Spira, Kevin R. Coombes, Ignacio I. Wistuba

Manuscript received March 11, 2013; revised December 13, 2013; accepted December 19, 2013.

Correspondence to: Ignacio I. Wistuba, MD, Departments of Translational Molecular Pathology and Thoracic/Head and Neck Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX (e-mail: iiwistuba@mdanderson.org).

Background Earlier work identified specific tumor-promoting abnormalities that are shared between lung cancers and adjacent normal bronchial epithelia. We sought to characterize the yet unknown global molecular and adjacent airway field cancerization (FC) in early-stage non–small cell lung cancer (NSCLC).

Methods Whole-transcriptome expression profiling of resected early-stage (I–IIIA) NSCLC specimens (n = 20) with matched tumors, multiple cytologically controlled normal airways with varying distances from tumors, and uninvolved normal lung tissues (n = 194 samples) was performed using the Affymetrix Human Gene 1.0 ST platform. Mixed-effects models were used to identify differentially expressed genes among groups. Ordinal regression analysis was performed to characterize site-dependent airway expression profiles. All statistical tests were two-sided, except where noted.

Results We identified differentially expressed gene features (n = 1661) between NSCLCs and airways compared with normal lung tissues, a subset of which (n = 299), after gene set enrichment analysis, statistically significantly ($P < .001$) distinguished large airways in lung cancer patients from airways in cancer-free smokers. In addition, we identified genes (n = 422) statistically significantly and progressively differentially expressed in airways by distance from tumors that were found to be congruently modulated between NSCLCs and normal lung tissues. Furthermore, *LAPTM4B,* with statistically significantly increased expression ($P < .05$) in airways with shorter distance from tumors, was upregulated in human immortalized cells compared with normal bronchial epithelial cells ($P < .001$) and promoted anchorage-dependent and -independent lung cancer cell growth.

Conclusions The adjacent airway FC comprises both site-independent profiles as well as gradient and localized airway expression patterns. Profiling of the airway FC may provide new insights into NSCLC oncogenesis and molecular tools for detection of the disease.

JNCI J Natl Cancer Inst (2014) 106(3): dju004 doi:10.1093/jnci/dju004

Earlier work by Slaughter et al. in patients with oral premalignant and cancer lesions suggested that histologically normal-appearing tissues adjacent to lesions display tumor-associated molecular abnormalities (1). Notably, Auerbach et al. demonstrated that cigarette smoke induces widespread histological changes and premalignant lesions in the bronchial epithelia in the lungs of smokers, suggestive of a field effect (2). This phenomenon, coined field cancerization (FC), was shown to be evident in various epithelial malignancies, including gastric, esophageal, hepatic, cervical, skin, and lung cancers (3–6) and was proposed to precede and explain the development of multiple primary and locally recurrent cancer (3,7).

Previously, an analysis of histologically normal epithelium and premalignant and malignant epithelia from lung squamous cell carcinoma (SCC) patients indicated that multiple, sequentially occurring allele-specific chromosomal deletions commence early in the multistage pathogenesis of SCCs (8,9). Notably, 31% of histologically normal epithelium specimens had clones of cells with allelic loss at one or more regions examined, including loss of heterozygosity at chromosomal regions 3p and 9p (8,10). Belinsky et al. identified promoter methylation of *p16*, a common aberration in lung tumors (11), in at least one bronchial epithelial site from 44% of lung cancer cases examined (12). In addition, our group and others have demonstrated that mutations in the epidermal growth factor receptor (*EGFR)* and *KRAS* oncogenes were also found in histologically normal tissue adjacent to lung tumors (13,14).

Global expression profiles have been described in bronchial epithelium of smokers, a portion of which exhibited cancer diagnostic properties (15–18), as well as in the field of injury of early-stage non–small cell lung cancer (NSCLC) patients who had their tumors surgically resected before airway transcriptome analysis

(19). However, the adjacent airway FC in NSCLC has not yet been characterized at a whole-transcriptome level. In this study, we performed expression profiling of matched NSCLCs, uninvolved normal lung tissue, and multiple airways with varying distances from tumors to define the transcriptomic architecture of the adjacent airway FC.

Methods

Lung Tumor Resected FC Specimens and Airway Epithelial Cell Collection

The FC specimens, comprised of lung tumors, uninvolved normal lung parenchyma, and multiple normal-appearing airways with varying distances from tumors, were obtained from early stage (I–IIIA) patients at MD Anderson Cancer Center. Tumor stage was classified as described previously (20). The study was approved by the institutional review boards, and all participants provided written informed consent. Malignant and paired normal lung tissues from each case patient were obtained snap-frozen, preserved in RNA*later* or by surface brushing. For each tissue sample, the percentage of malignant tissue was calculated by histological examination (J. Fujimoto) after hematoxylin and eosin staining. All malignant samples contained more than 40% tumor cells. Twenty NSCLC FC case patients were included in the study and, along with their clinicopathological information, are summarized in Supplementary Table 1 (available online).

Airway epithelia were obtained by brushing three to five sequential bronchiolar structures with varying distances from tumors (Supplementary Figure 1, available online) using sterile Cytosoft cytology brushes (Medical Packaging Corporation, Camarillo, CA). The spatial distance between two consecutive airway brushings was similar (approximately 2 cm). Airways were denoted by numbers 1 (relatively closest from tumor) to 5 (relatively farthest). The relative distance of an airway brushing (eg, airway 1) from the adjacent NSCLC tumor was similar across all case patients. Airway brushings were placed in Qiazol lysis buffer (Qiagen, Valencia, CA) in dry ice and immediately stored at –80°C. Confirmation of epithelial cell collection by pan-cytokeratin immunohistochemical analysis, as well as lack of neoplastic or preneoplastic cells (Supplementary Figure 1, available online), was performed as described in the Supplementary Methods (available online).

Microarray Data Analysis

RNA samples were processed for microarray expression profiling using the Affymetrix Human Gene 1.0 ST platform (Affymetrix, Santa Clara, CA) (Supplementary Methods, available online). Raw data were quantified using background correction, quantile normalization, and robust multichip analysis (21) probe-level models and summarization methods. Minimum Information About a Microarray Experiment (MIAME)-compliant data were submitted to the Gene Expression Omnibus under series GSE44077 (samples GSM1077844–GSM1078069). Basic quality control was assessed by graphical summaries of array intensities and Bland-Altman (M vs A) plots (Supplementary Methods, Sweave Report 1, available online). Linear mixed-effects models were used to characterize gene features concordantly differentially expressed (in the same direction) between both NSCLCs and airways compared with normal lung tissues (site-independent analysis). The different groups were set as fixed effects, whereas patients were set as random factors in the mixed-effects models. An ordinal logistic regression model (22–24) was applied to identify gene features that were statistically significantly decreased or increased in airways by distance from tumors (site-dependent analysis). The model assumes that each FC case comprises only one airway brushing at a specific spatial location (eg, airway 1) and fits expression patterns that change in a gradient fashion with different airway spatial locations (airways 1–5) classified as categorical ordered labels. One-sided *t* tests were used to determine whether site-dependent differential airway expression exhibits similar directional patterns between NSCLCs and matched normal lung tissues. A site-dependent FC score was generated to collectively signify and plot gene features that are differentially modulated with respect to proximity from tumors across all airways by distance from tumors (airways 1–5) and between paired NSCLCs and uninvolved normal lung tissues. The score was calculated by the sum of gene features upregulated in airways with shorter distance from tumors minus the sum of gene features downregulated in airways with shorter distance from tumors. To adjust for multiple testing in all analyses, beta-uniform mixture models were used to estimate false discovery rates as previously described (25). Pathways analysis was performed using ingenuity pathways analysis.

Statistical Analysis

Fisher exact test was used to determine statistical significance of the difference in airway type (cancer vs no cancer) between clusters after hierarchical clustering analysis. Kruskal–Wallis test was used to test for statistical significance of differences in quantitative real-time polymerase chain reaction (QRTPCR)–based expression of genes among NSCLCs, airways, and normal lung tissues. Analysis of variance was used to test for statistical significance of differences in QRTPCR-based expression of genes among airways with respect to proximity from tumors. Student *t* test was used to assess for statistical significance among different groups in the in vitro experiments. All statistical tests were two-sided, except where noted. Additional methods including all details and R codes of the microarray statistical analysis are included in the Supplementary Methods and in the Supplementary Sweave Reports (available online).

Results

Identification of Adjacent Airway FC Profiles in Early-Stage NSCLC

The FC in the airway adjacent to NSCLC has not been characterized at the global transcriptomic level before. We analyzed the transcriptomes of cytologically controlled NSCLC tumors, paired uninvolved normal lung tissue, and brushings of normal airway epithelia collected at sequentially varying distances from the tumors (n = 194 samples) (Supplementary Figure 1). Samples were obtained from specimens surgically resected from 20 patients (n = 9 women and 11 men; n = 5 never-smokers and 15 smokers) with stages I to III NSCLC (n = 14 adenocarcinomas, 5 SCCs, and 1 not otherwise specified NOS NSCLC) (Supplementary Table 1, available online).

A schematic of the study's design and different analyses is represented in Figure 1. We first sought to characterize global FC

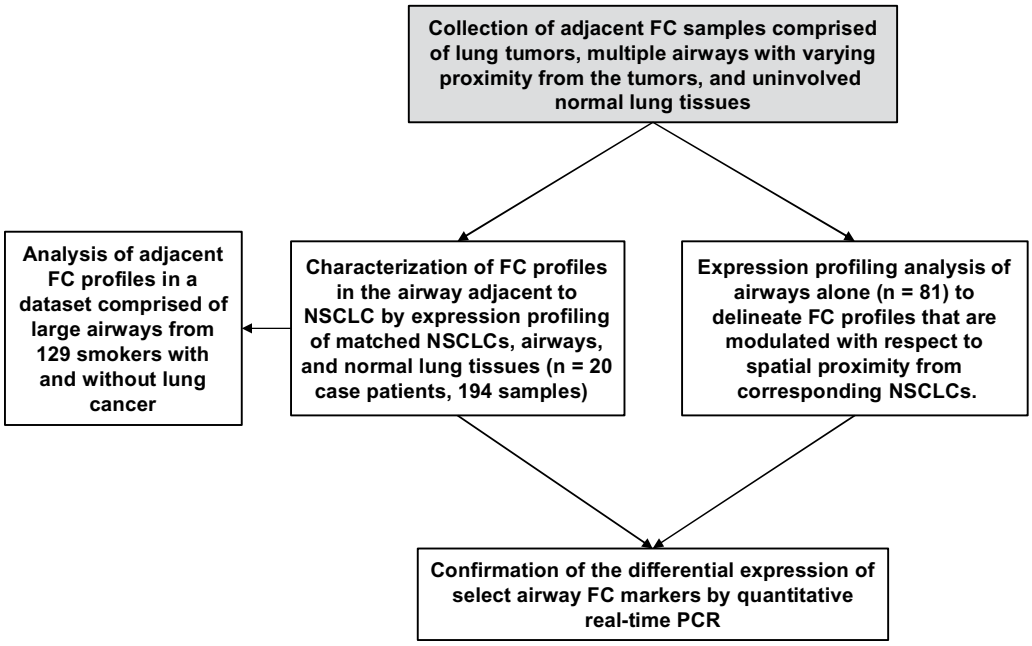


Figure 1. Schematic of the transcriptomic analysis of the airway field cancerization (FC) in non–small cell lung cancer (NSCLC). FC samples comprised of matched lung tumors, multiple cytologically controlled normal airways with varying proximity from tumors, and uninvolved normal lung tissues were obtained from 20 resected early-stage (I–IIIA) NSCLC specimens. Gene expression profiling of all the samples (n = 194) was performed to characterize global FC profiles (differentially expressed in the same direction between both tumors and airways compared with normal lung tissues) in the normal-appearing airway adjacent to NSCLCs (site-independent analysis). Adjacent FC profiles were then analyzed by gene set enrichment analysis (GSEA) in a set of large airways of smokers (n = 129) with and without lung cancer (17) to identify FC profiles that can distinguish lung cancer among smokers. Global expression profiling was also used to delineate FC profiles that are modulated in the airway with respect to proximity from corresponding NSCLCs (site-dependent analysis). Quantitative real-time polymerase chain reaction (PCR) was used to confirm the differential expression of select airway FC markers.

profiles in the normal-appearing airway adjacent to the NSCLC. We identified 1661 gene features (n = 457 upregulated and 1204 downregulated) (Supplementary Table 2, available online) that were statistically significantly and concordantly differentially expressed between both NSCLCs and airways compared with normal lung tissue (false discovery rate–corrected $P < .01$; fold-change > 1.5) (Figure 2A; Supplementary Methods, Sweave Report 2, available online). Further analysis identified statistically significantly differentially expressed gene features between the airway of adenocarcinomas and SCCs (n = 415) and between the airway of never-smoker and smoker NSCLCs (n = 51) (Supplementary Methods, Sweave Report 2, available online). Pathway analysis of the 1661 gene features by ingenuity pathways analysis revealed modulation of key cancer-associated pathways and gene-interaction networks (all $P < .05$) (Figure 2B). The most statistically significantly ($P < .001$) modulated pathway and the gene network with highest number of differentially expressed and closely related (G-protein coupled receptors) interacting genes are represented in Figure 2, C and D, respectively. These findings highlight expression patterns and pathways that are typically deregulated in overt tumors but are also prevalent in histologically normal airway epithelia.

Analysis of Adjacent Airway FC Profiles in Smoker Patients With Suspected Lung Cancer

We then sought to ascertain whether the adjacent FC profiles may be indicative of lung cancer among smokers. We examined the expression of the adjacent airway FC profile in a cohort from Spira et al. (17) comprised of 129 large airway samples from smokers with and without lung cancer. After matching gene features between the two studies by common Entrez Gene identifiers, 261 upregulated and 749 downregulated adjacent airway FC gene features remained, which were used to perform gene set enrichment analysis (GSEA) with identifiers ranked according to the Student t statistic between smokers with and without lung cancer. This analysis demonstrated that the gene features that were downregulated in the adjacent airway FC were statistically significantly ($P < .001$) enriched within gene features downregulated in large airways of smokers with lung cancer compared with airways of cancer-free smokers, although such a statistically significant connection was not observed for the upregulated genes. Leading edge gene sets comprised of 299 genes (Supplementary Table 3, available online) that were concordantly modulated between NSCLCs and airways compared with normal lung tissue in the adjacent airway FC and between large airways of smokers with and without lung cancer ($P < .001$) (Figure 3) were then derived [as described previously (26)]. These data suggest that the adjacent airway FC harbors potential markers for detection of lung cancer among smokers.

Site-Dependent Analysis of the Adjacent Airway FC by Distance From Tumors

We then sought to determine whether the adjacent airway FC transcriptome varies with respect to tumor proximity. We performed ordinal logistic regression analysis of the transcriptomes of airway samples (n = 81) obtained at varying distances from the tumors. Using a 5% false discovery rate, we identified 422 gene features (n = 335 upregulated and 87 downregulated with shorter distance

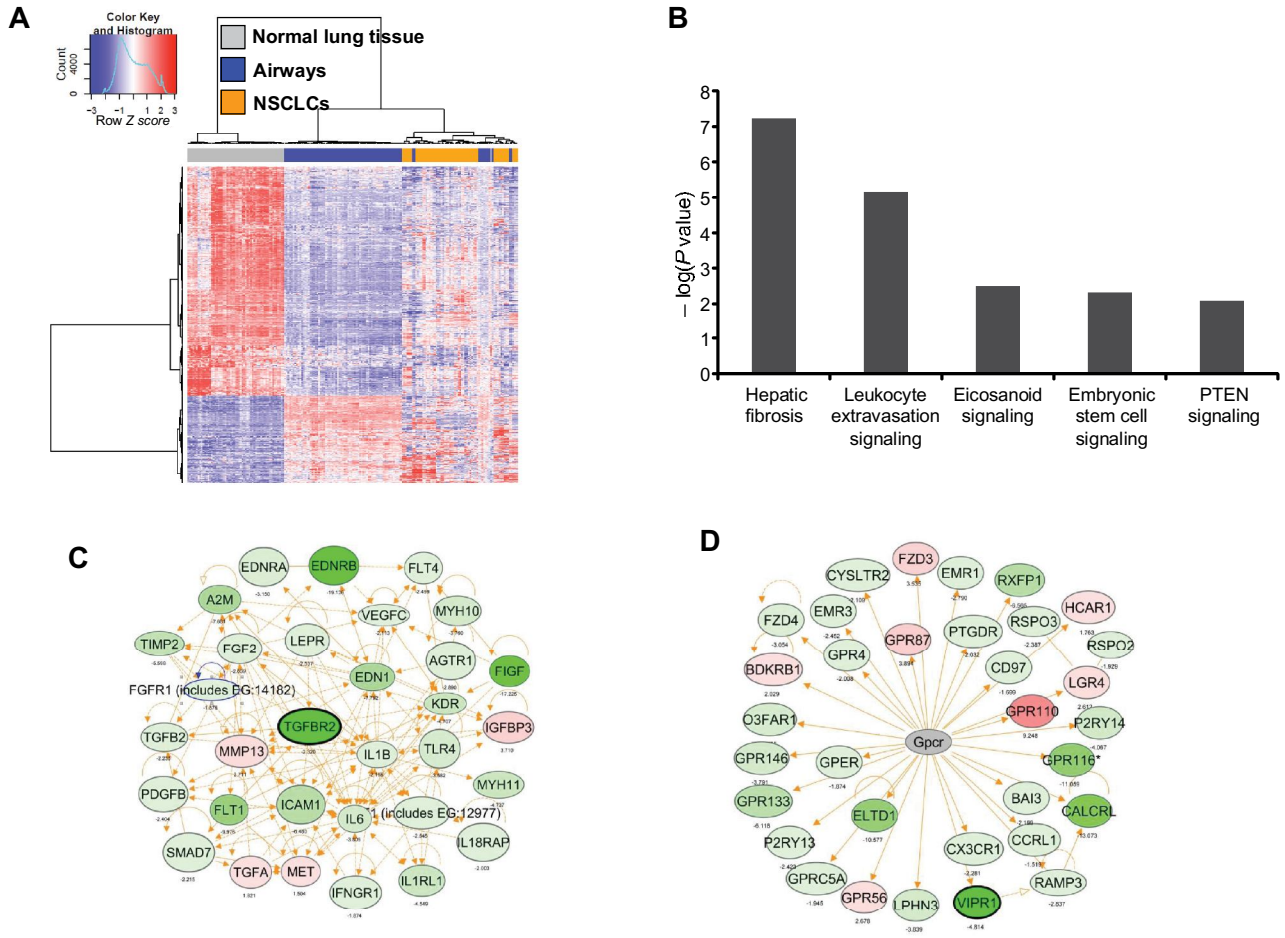


Figure 2. Identification of adjacent airway field cancerization (FC) profiles in non–small cell lung cancer (NSCLC). **A**) Hierarchical clustering of gene features statistically significantly differentially expressed between both NSCLCs and airways compared with normal lung tissues (n = 1661). **Columns** represent samples (n = 194 samples from 20 case patients), and **rows** represent gene features (**red** = upregulated; **blue** = downregulated). **B**) Functional pathways analysis using ingenuity pathways analysis (IPA) of the differentially expressed genes. Statistical significance of the identified overrepresented canonical pathways is indicated by the negative log of the *P* values. Functional pathways and interaction network analysis by IPA depicting the most statistically significantly ($P < .001$) modulated pathway (**C**) and the gene network with highest number of differentially expressed and closely related (G-protein coupled receptors) interacting genes (**D**); **red** = higher experession; **green** = lower expression. Genes selected for subsequent confirmation by quantitative real-time polymerase chain reaction are highlighted by **black margins**.

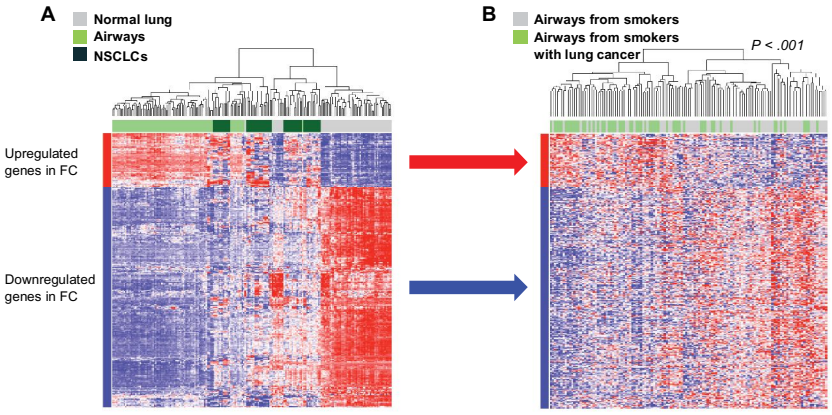


Figure 3. Analysis of adjacent airway field cancerization (FC) profiles in large airways of patients with suspected lung cancer. Gene set enrichment analysis (GSEA) was performed, as described in the Supplementary Methods (available online), to identify genes that were upregulated and downregulated in the adjacent airway FC that were also concordantly enriched between genes differentially expressed between histologically normal airways of smokers with and without lung cancer (17). Hierarchical cluster analysis using the 299 leading edge genes (n = 59 upregulated and 240 downregulated) was performed side by side in the adjacent airway FC (**A**) and in airways of smokers with and without lung cancer (**B**). **Columns** represent samples, and **rows** represent gene features (**red** = upregulated; **blue** = downregulated). *P* value showing statistical significance of separate clustering of airways of smokers with lung cancer from those of healthy smokers was obtained by the two-sided Fisher exact test.

from the tumors) (Supplementary Table 4 and Supplementary Methods, Sweave Report 3, available online) and key cancer-associated signaling pathways that were differentially expressed in airways with respect to tumor proximity (Figure 4, A and B). We then derived a quantitative score (Supplementary Methods, available online) to signify the extent of the site-dependent effect (Figure 4C). It is worthwhile to mention that the site-dependent effect was more pronounced in lung SCCs than in adenocarcinomas (Supplementary Figure 2, A and B, available online).

We then examined whether the 422 gene features were modulated concordantly between NSCLCs and normal lung tissues. We performed one-sided *t* tests of the gene features between NSCLC and normal lung tissues to identify those that are modulated in the same direction between the tumors and uninvolved normal lung tissues. We found that 291 of the 335 genes that were increased and 53 of the 87 that were decreased in airways with shorter distance from tumors were also upregulated and downregulated, respectively, in NSCLCs compared with normal lung tissues (Supplementary Table 5 and Supplementary Methods, Sweave Report 4, available online). In addition, the site-dependent airway FC score was statistically significantly and concordantly modulated between NSCLCs and paired uninvolved normal lung tissues (Figure 4D). These findings suggest that the molecular airway FC in NSCLC is in part localized and modulated by distance from tumors and that this gradient site-dependent effect in the airway recapitulates NSCLC expression patterns.

QRTPCR Analysis of the Differential Expression of Adjacent Airway FC Markers

We performed QRTPCR analysis of the expression of transforming growth factor beta receptor II (*TGFBR2*) and vasoactive intestinal peptide receptor 1 (*VIPR1*), which were selected based on pathway-based analysis of the site-independent adjacent airway FC profile (Figure 2, C and D) and of neuropilin (NRP) and tolloid (TLL)-like 2 (*NETO2*) and lysosomal protein transmembrane 4 beta (*LAPTM4B*), which were among the FC markers (Supplementary Table 4, available online) with

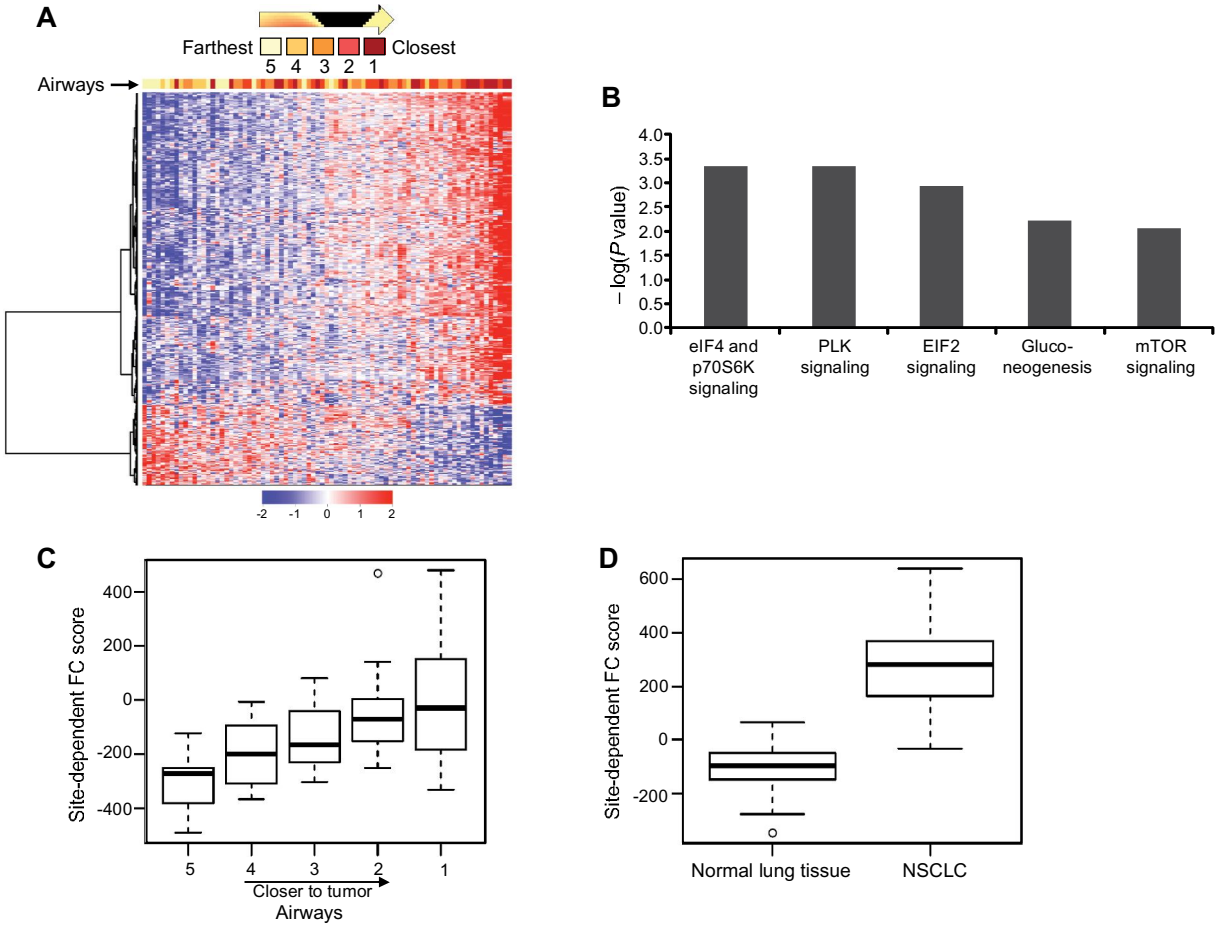


Figure 4. Analysis of airway expression profiles by distance from corresponding non–small cell lung cancers (NSCLCs). Ordinal logistic regression analysis of airways was performed, as described in the Supplementary Methods (available online), and identified 422 gene features with statistically differential expression in the airway with respect to tumor proximity (false discovery rate < 5%). **A**) Clustering analysis was performed as described in the Supplementary Sweave Reports (available online), and airway samples were arranged by difference in the expression values of the site-dependent genes between the upregulated and downregulated gene clusters. **Columns** represent samples, and **rows** represent gene features (**red** = upregulated; **blue** = downregulated). **B**) Functional pathways analysis by ingenuity pathways analysis of the site-dependent differentially expressed genes. Statistical significance of the identified overrepresented canonical pathways is indicated by the negative log of the *P* values. The site-dependent effect in the adjacent field cancerization (FC) was quantified as described in the Supplementary Methods and Sweave Reports (available online). Box plots depicting site-dependent FC score in airways (**C**) and between corresponding paired NSCLCs and normal lung tissue after statistical analysis by one-sided *t* tests (**D**). **Heavy lines** indicate medians, and **whiskers** indicate maximum and minimum FC scores. Airway distance from tumors is numerically indicated with a range of 1 (closest) to 5 (farthest).

statistically significantly ($P < .01$) increased expression in airways with shorter distance from tumors. *VIPR1* also exhibited statistically significant site-dependent expression modulation in the FC (Supplementary Table 4, available online). QRTPCR demonstrated that *TGFBR2* and *VIPR1* mRNA levels were statistically significantly decreased and *NETO2* and *LAPTM4B* levels increased in NSCLCs and airways compared with normal lung tissue ($P < .001$) (Figure 5, A–D, top panels). QRTPCR also revealed that *VIPR1* ($P = .02$) but not *TGFBR2* exhibited statistically significantly decreased expression in airways with respect to tumor proximity (Figure 5, A and B, middle panels), whereas *NETO2* and *LAPTM4B* levels were statistically significantly increased in the airway with shorter distance from tumors (both $P < .05$) (Figure 5, C and D, middle panels). Microarray and QRTPCR-based expression of the four genes were statistically correlated ($P < .001$) (Figure 5, A–D, lower panels). These findings highlight the confirmed differential expression of markers in the adjacent airway FC.

Effect of the FC Gene *LAPTM4B* on Lung Cancer Cell Growth

We were prompted to study the relevance of *LAPTM4B*, a lysosome-associated transmembrane putative oncogene (27) with no known role in lung carcinogenesis, to the lung malignant phenotype. We first found that *LAPTM4B* expression was statistically significantly increased in BEAS-2B immortalized lung epithelial cells compared with normal bronchial cells ($P < .001$) (Supplementary Figure 3, available online) and in NSCLC cell lines (data not shown) compared with bronchial epithelial cells after in silico analysis of a publicly available dataset of human NSCLC and bronchial epithelial cell lines [Gene Expression Omnibus dataset GSE4824 (28)]. Transient knockdown of *LAPTM4B* expression in immortalized and malignant lung epithelial cell lines effectively and statistically significantly reduced *LAPTM4B* expression (all $P < .001$) and cell growth (all $P < .05$) (Supplementary Figures 4 and 5, available online). Stable knockdown of *LAPTM4B* in Calu-6 lung cancer cells statistically significantly suppressed *LAPTM4B* expression (relative expression compared with empty vector, mean ± standard deviation [SD]: sh*LAPTM4B* clone 1: 0.21 ± 0.02; sh*LAPTM4B* clone 2: 0.31 ± 0.03; $P < .001$) (Figure 6A) concomitant with statistically significantly reduced cell growth (relative cell numbers × 10^4 compared to 0 hours, mean ± SD: empty vector: 5.83 ± 0.38; scrambled shRNA: 5.69 ± 0.75; sh*LAPTM4B* clone 1: 3.72 ± 0.19, $P = .01$; sh*LAPTM4B* clone 2: 3.67 ± 0.60, $P = .02$) (Figure 6B) and anchorage-dependent (mean colony numbers ± SD: empty vector:

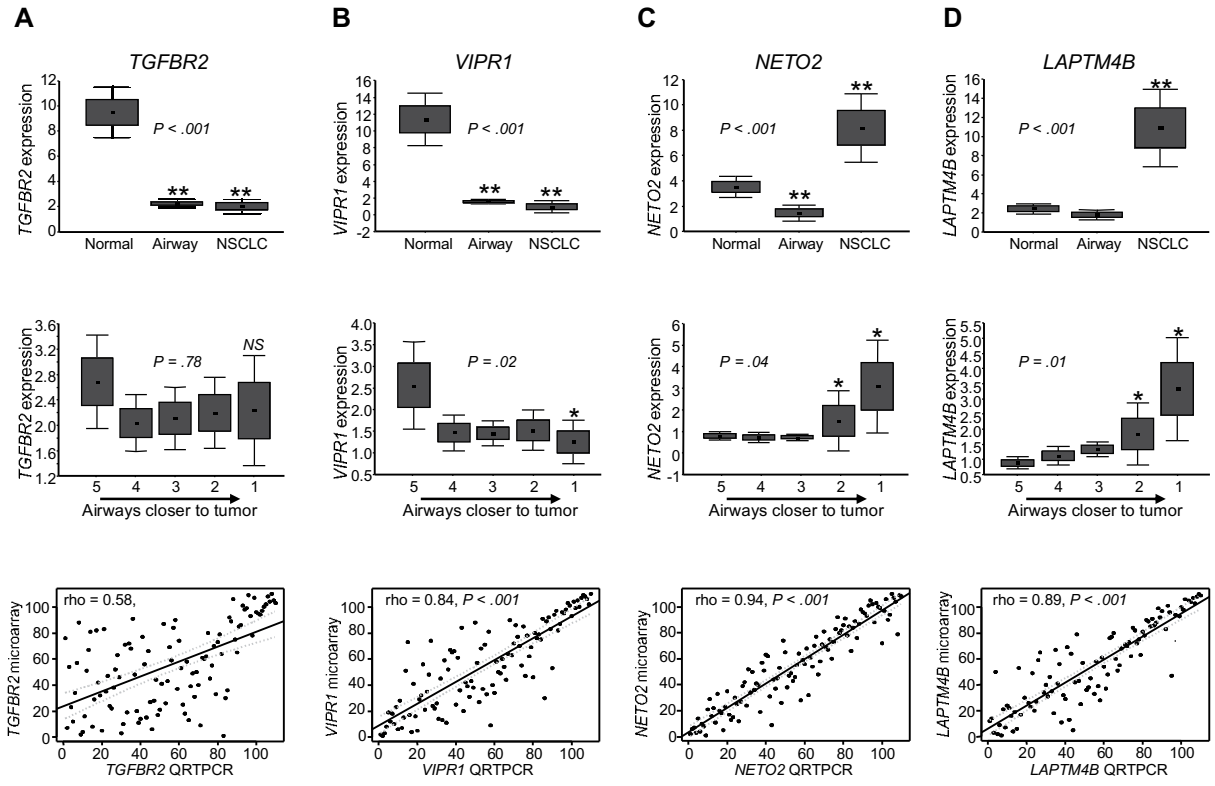


Figure 5. Quantitative real-time polymerase chain reaction (QRTPCR) analysis of airway field cancerization (FC) markers. Expression of *TGFBR2* (**A**), *VIPR1* (**B**), *NETO2* (**C**), and *LAPTM4B* (**D**) was analyzed by QRTPCR in 18 of 20 NSCLC FC case subjects studied with sufficient RNA from airway samples left over after expression profiling. Expression of the indicated genes is depicted by group (NSCLCs, airways, and normal lung tissues; **upper panels**) and across airway samples based on distance from corresponding NSCLCs (1 = airway closest to tumors; 5 = airways relatively farthest from tumors; **middle panels**). Relative mRNA expression was assessed by QRTPCR, normalized to that of *TBP*, and quantified using the $2^{-\Delta\Delta CT}$ relative quantification method as detailed in the Supplementary Methods (available online). PCR reactions for each FC sample were carried out in duplicate. **Boxes** indicate ± standard error of the mean; **error bars** indicate standard deviation. Statistical significance of differences in expression among NSCLC, airway, and normal lung groups was assessed by the Kruskal–Wallis test (**upper panels**), and statistical significance of differences among different airways was assessed by analysis of variance (**middle lanes**). Correlation between expression of the indicated genes quantified by microarray and QRTPCR analyses was statistically assessed by Spearman rank (**lower panels**). *LAPTM4B* = lysosomal protein transmembrane 4 beta; *NETO2* = neuropilin (NRP) and tolloid (TLL)-like 2; *TGFBR2* = transforming growth factor beta receptor II; *VIPR1* = vasoactive intestinal peptide receptor 1; TBP = TATA box binding protein. **P* < .05; ** *P* < .001; NS, not significant. All statistical tests were two-sided.

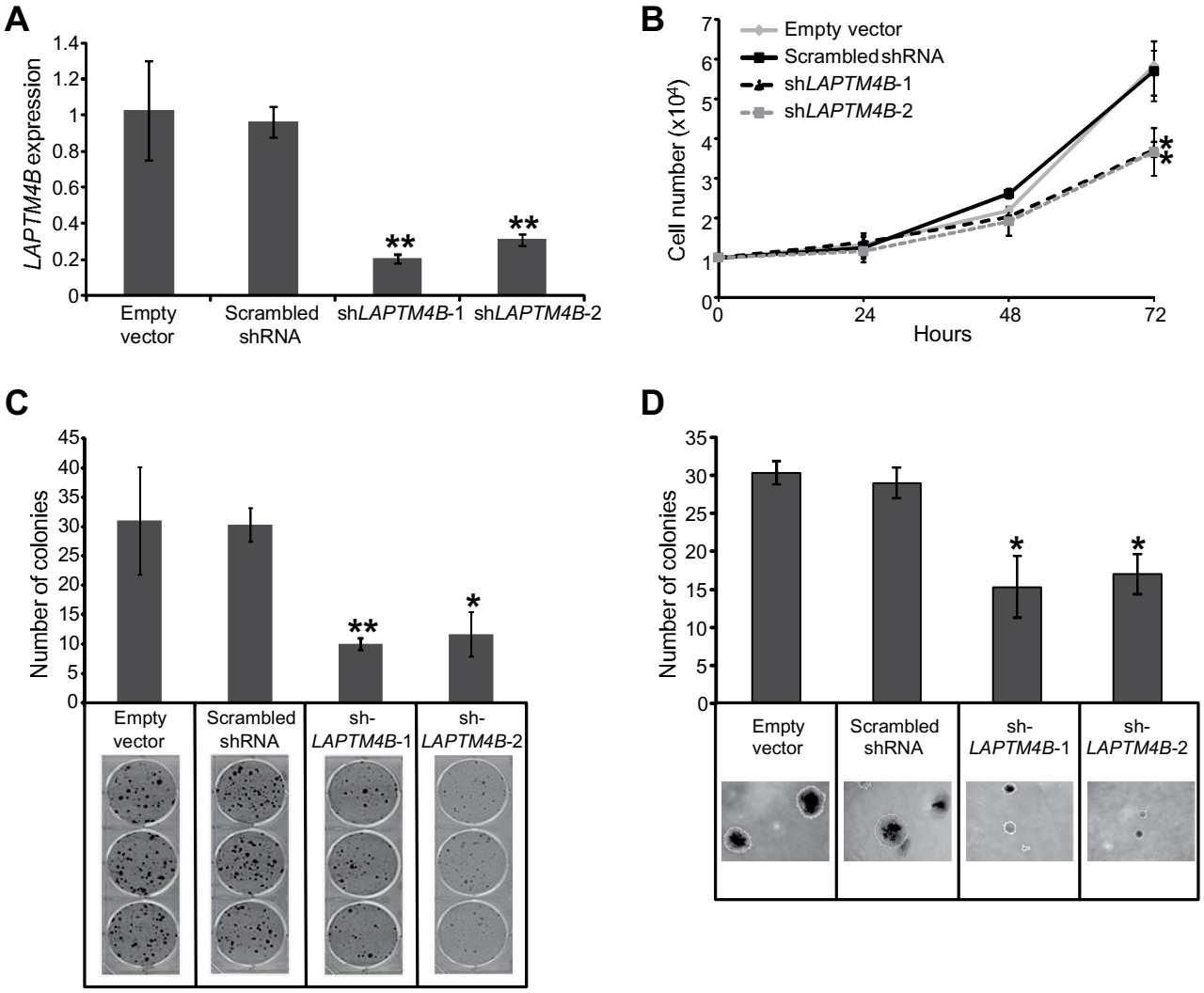


Figure 6. Effect of RNA interference-mediated knockdown of *LAPTM4B* on lung cancer cell anchorage-dependent and -independent growth and colony formation. Calu-6 lung cancer cells were stably transfected with empty vectors, vectors containing scrambled short hairpin RNA (shRNA), as well as vectors with *LAPTM4B*-specific shRNA sequences as described in the Methods section. **A**) Quantitative real-time polymerase chain reaction analysis depicting statistically significantly reduced *LAPTM4B* relative expression in two Calu-6 sublines (sh*LAPTM4B*-1 and sh*LAPTM4B*-2) stably transfected with two different *LAPTM4B*-specific shRNA sequences compared with Calu-6 cells stably transfected with empty and scrambled shRNA-containing vectors. **B**)The indicated stably transfected Calu-6 cells were seeded in triplicates (5×10^4 cells/well) in 12-well plates for 72 hours, after which the number of cells in each well was computed by the trypan blue exclusion method. Cells were seeded in triplicates at a seeding density of 250 cells per well in six-well culture plates or 150 cells per well on soft agar for assessment of anchorage-dependent (**C**) and -independent (**D**) colony formation, respectively. Cell colonies were then quantified as described in the Methods section. All experiments were done in triplicate. Representative images of cell colonies are depicted in the **lower panels** and were obtained with a phase-light microscope. **Error bars** indicate standard deviation. $*P < .05$; $**P < .001$ by the two-sided Student *t* test.

31.0±9.17; scrambled shRNA: 30.3±2.89; sh*LAPTM4B* clone 1: 10.0±1.0, $P < .001$; sh*LAPTM4B* clone 2: 11.67±3.79, $P = .002$) (Figure 6C) and -independent colony number (mean colony numbers ± SD: empty vector: 30.3±1.53; scrambled shRNA: 29.0±2.0; sh*LAPTM4B* clone 1: 15.3±4.04, $P = .003$; sh*LAPTM4B* clone 2: 17.0±2.65, $P = .006$) (Figure 6D). These data demonstrate that *LAPTM4B* is a positive mediator of immortalized and malignant lung epithelial cell growth.

Discussion

In this study we characterized the airway FC transcriptome adjacent to NSCLC. We identified gene features that were statistically significantly and concordantly modulated between both NSCLCs and airways compared with normal lung tissue, a subset of which was indicative of lung cancer among smokers. Moreover, we revealed that the adjacent airway FC exhibits gradient site-dependent expression patterns with respect to tumor proximity, which effectively predicted NSCLC profiles, pointing to their possible roles in lung cancer pathogenesis. In addition, *LAPTM4B*, whose expression was increased in the airway with shorter distance from the tumor, was elevated in NSCLC and in immortalized lung epithelial cells and promoted anchorage-dependent and -independent lung cancer cell growth in vitro.

Our analyses pointed to the statistically significant differential expression of FC markers—namely, *TGFBR2*, *VIPR1*, *NETO2* and

LAPTM4B. *TGFBR2*, a transmembrane receptor serine threonine kinase that mediates TGF-β signaling (29), was previously reported to be downregulated in invasive adenocarcinomas compared with bronchioalveolar carcinomas (30), and loss of the murine counterpart was shown to mediate progression and development of invasive adenocarcinomas in a mouse model of *Kras*-induced lung cancer (31). *VIPR1* mRNA expression was shown to be highest in normal lung tissue compared with various human normal epithelial tissues examined and to peripheral blood leukocytes (32). *NETO2* expression was reported to be upregulated in proliferating hemangiomas (33) and associated with invasiveness and motility of cancer cells (34). Furthermore, *LAPTM4B* was shown to mediate prosurvival autophagy and chemoresistance in breast tumor cells (35) and was found to be upregulated and associated with poor prognosis in ovarian and hepatocellular carcinomas (36,37). In this study we showed that *LAPTM4B*, whose role in lung cancer was previously unknown, was upregulated in NSCLCs and promoted anchorage-dependent and -independent lung cancer cell growth. Our findings suggest that detailed interrogation of the airway FC may be a useful approach to highlight potential uncharacterized mechanisms and molecules involved in lung cancer pathogenesis.

Our group has recently portrayed the spatial and temporal molecular field of injury in early-stage NSCLC patients by expression profiling of large airways after definitive surgery (19). It is important to mention that in our previous study, normal airway epithelia were collected by endoscopic bronchoscopy brushings within 12 months after surgical removal of the tumor and when NSCLC tumors were not present in situ at time of collection (19). In our current study, we performed expression profiling of multiple normal-appearing airways at various distances from tumors in conjunction with paired NSCLCs and normal lung tissues that were still in situ at the time of airway epithelia collection. It is important to note that although the identified adjacent airway FC profiles were, in part, enriched in large proximal airways, they were substantially dissimilar in the normal lung tissue samples (Figures 2 and 3). It is reasonable to suggest that the identified field effects are more readily discerned in samples (eg, airways) with a higher fraction of epithelial cell content and raises the notion that some of the changes in the adjacent airway FC may be related to cell type. However, this notion is also applicable to the comparison of NSCLCs with normal lung tissue, an analysis that is commonly performed. In addition and through analysis of normal airways alone, we identified gradient and localized site-dependent expression profiles within the adjacent airway FC that predicted NSCLC profiles. It is intriguing to presume that gradient site-dependent airway field effects may be associated with the development of NSCLC tumors in a particular lobe or anatomical region of the lung. However, it cannot be discerned whether these airway FC effects are a cause or consequence of NSCLC development, although it is not unlikely that they are involved in lung carcinogenesis, as embodied by the effect of the site-dependent airway FC marker, *LAPTM4B*, on lung cancer cell growth. This supposition can be addressed in future studies by assessment of the FC in lung cancer patients before and after surgery.

Analysis of a dataset of airways from smokers with suspected lung cancer (17) revealed that a subset of genes in the adjacent airway FC profile was able to distinguish lung cancer patients among smokers with suspicion of the disease. It is important to mention, however, that the airway FC markers we had identified to be modulated with respect to spatial proximity from tumors were not able to statistically significantly identify lung cancer among smokers. This observation may be attributable to the possible strong association of the gradient site-dependent profiles with the compartments that are adjacent or local to the tumor (6,10). It is plausible to surmise that the adjacent airway FC harbors molecular cancerization profiles that are clinically relevant to both the detection and (chemo) prevention of lung cancer and those that are biologically relevant to understanding the early pathogenesis of this malignancy.

It is worthwhile to mention that our study is not without limitations. Further analysis to fully characterize how the adjacent airway FC varies by NSCLC histology (squamous vs nonsquamous) and smoking status (never-smoker vs smoker) was hindered by the limited number of FC case patients. Moreover, although we compared and contrasted the gradient and site-dependent airway FC profile (n = 422 gene features) that we had derived from NSCLC case patients, among different subgroups (eg, by histology), identification of site-dependent profiles unique to a histological or smoking subtype of NSCLC was also impeded by the small number of FC case patients. Additionally, we could not connect, at the present time, adjacent FC profiles of nonsmoker adenocarcinoma patients to the field effect in the large airway because of the paucity of such airway samples from nonsmoker lung cancer patients. Furthermore, it is reasonable to speculate that RNA sequencing (38), compared with gene expression profiling technology that we used, would provide a more thorough characterization (eg, identification of novel tissue-specific transcripts) of the transcriptomic architecture of the airway FC. Nonetheless, our study represents the first attempt to characterize the global adjacent airway FC in NSCLC, and efforts are underway to expand this working model into different subtypes of lung cancer, including never-smoker adenocarcinomas, and to apply more advanced technologies and platforms (eg, RNA sequencing) for studying the airway FC.

In conclusion, our gene expression profiling efforts revealed that the adjacent and molecular airway FC in NSCLC is comprised of markers that can identify lung cancer among smokers as well as gradient and localized site-dependent expression patterns that recapitulate NSCLC profiles. Our findings suggest that profiling of the airway FC in conjunction with NSCLCs may provide additional insights into the biology of NSCLC and the development of molecular tools for the detection of the malignancy.

References

1. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer*. 1953;6(5):963–968.
2. Auerbach O, Stout AP, Hammond EC, et al. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. *N Engl J Med*. 1961;265(6):253–267.
3. Braakhuis BJ, Tabor MP, Kummer JA, et al. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. *Cancer Res*. 2003;63(8):1727–1730.
4. Gomperts BN, Spira A, Massion PP, et al. Evolving concepts in lung carcinogenesis. *Semin Respir Crit Care Med*. 2011;32(1):32–43.
5. Graham TA, McDonald SA, Wright NA. Field cancerization in the GI tract. *Future Oncol*. 2011;7(8):981–993.

6. Kadara H, Wistuba II. Field cancerization in non-small cell lung cancer: implications in disease pathogenesis. *Proc Am Thorac Soc*. 2012;9(2):38–42.
7. Gold KA, Kim ES, Lee JJ, et al. The BATTLE to personalize lung cancer prevention through reverse migration. *Cancer Prev Res (Phila)*. 2011;4(7):962–972.
8. Wistuba, II, Behrens C, Milchgrub S, et al. Sequential molecular abnormalities are involved in the multistage development of squamous cell lung carcinoma. *Oncogene*. 1999;18(3):643–650.
9. Wistuba II, Behrens C, Virmani AK, et al. High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. *Cancer Res*. 2000;60(7):1949–1960.
10. Wistuba, II, Gazdar AF. Lung cancer preneoplasia. *Annu Rev Pathol*. 2006;1:331–348.
11. Belinsky SA, Nikula KJ, Palmisano WA, et al. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc Natl Acad Sci U S A*. 1998;95(20):11891–11896.
12. Belinsky SA, Palmisano WA, Gilliland FD, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res*. 2002;62(8):2370–2377.
13. Nelson MA, Wymer J, Clements N Jr. Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers. *Cancer Lett*. 1996;103(1):115–121.
14. Tang X, Shigematsu H, Bekele BN, et al. EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. *Cancer Res*. 2005;65(17):7568–7572.
15. Beane J, Sebastiani P, Liu G, et al. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol*. 2007;8(9):R201.
16. Spira A, Beane J, Shah V, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci U S A*. 2004;101(27):10143–10148.
17. Spira A, Beane JE, Shah V, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med*. 2007;13(3):361–366.
18. Gustafson AM, Soldi R, Anderlind C, et al. Airway PI3K pathway activation is an early and reversible event in lung cancer development. *Sci Transl Med*. 2010;2(26):26ra25.
19. Kadara H, Shen L, Fujimoto J, et al. Characterizing the molecular spatial and temporal field of injury in early-stage smoker non-small cell lung cancer patients after definitive surgery by expression profiling. *Cancer Prev Res (Phila)*. 2013;6(1):8–17.
20. Behrens C, Lin HY, Lee JJ, et al. Immunohistochemical expression of basic fibroblast growth factor and fibroblast growth factor receptors 1 and 2 in the pathogenesis of lung cancer. *Clin Cancer Res*. 2008;14(19):6014–6022.
21. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4(2):249–264.
22. Bull SB. Sample size and power determination for a binary outcome and an ordinal exposure when logistic regression analysis is planned. *Am J Epidemiol*. 1993;137(6):676–684.
23. Molaei M, Mansoori BK, Mashayekhi R, et al. Mucins in neoplastic spectrum of colorectal polyps: can they provide predictions? *BMC Cancer*. 2010;10(October 2010):537.
24. Perez-Ortiz M, Gutierrez PA, Hervas-Martinez C. Projection-based ensemble learning for ordinal regression. *IEEE Trans Cybern*. 2013; doi:10.1109/TCYB.2013.2266336.
25. Pounds S, Morris SW. Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics*. 2003;19(10):1236–1242.
26. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.
27. Li L, Wei XH, Pan YP, et al. LAPTM4B: a novel cancer-associated gene motivates multidrug resistance through efflux and activating PI3K/AKT signaling. *Oncogene*. 2010;29(43):5785–5795.
28. Zhou BB, Peyton M, He B, et al. Targeting ADAM-mediated ligand cleavage to inhibit HER3 and EGFR pathways in non-small cell lung cancer. *Cancer Cell*. 2006;10(1):39–50.
29. Kaklamani VG, Pasche B. Role of TGF-beta in cancer and the potential for therapy and prevention. *Expert Rev Anticancer Ther*. 2004;4(4):649–661.
30. Borczuk AC, Kim HK, Yegen HA, et al. Lung adenocarcinoma global profiling identifies type II transforming growth factor-beta receptor as a repressor of invasiveness. *Am J Respir Crit Care Med*. 2005;172(6):729–737.
31. Borczuk AC, Sole M, Lu P, et al. Progression of human bronchioloalveolar carcinoma to invasive adenocarcinoma is modeled in a transgenic mouse model of K-ras-induced lung cancer by loss of the TGF-beta type II receptor. *Cancer Res*. 2011;71(21):6665–6675.
32. Sreedharan SP, Huang JX, Cheung MC, et al. Structure, expression, and chromosomal localization of the type I human vasoactive intestinal peptide receptor gene. *Proc Natl Acad Sci U S A*. 1995;92(7):2939–2943.
33. Calicchio ML, Collins T, Kozakewich HP. Identification of signaling systems in proliferating and involuting phase infantile hemangiomas by genome-wide transcriptional profiling. *Am J Pathol*. 2009;174(5):1638–1649.
34. Horak CE, Lee JH, Elkahloun AG, et al. Nm23-H1 suppresses tumor cell motility by down-regulating the lysophosphatidic acid receptor EDG2. *Cancer Res*. 2007;67(15):7238–7246.
35. Li Y, Zhang Q, Tian R, et al. Lysosomal transmembrane protein LAPTM4B promotes autophagy and tolerance to metabolic stress in cancer cells. *Cancer Res*. 2011;71(24):7481–7489.
36. Yang H, Xiong F, Qi R, et al. LAPTM4B-35 is a novel prognostic factor of hepatocellular carcinoma. *J Surg Oncol*. 2010;101(5):363–369.
37. Yin M, Li C, Li X, et al. Over-expression of LAPTM4B is associated with poor prognosis and chemotherapy resistance in stages III and IV epithelial ovarian cancer. *J Surg Oncol*. 2011;104(1):29–36.
38. Mutz KO, Heilkenbrinker A, Lonne M, et al. Transcriptome analysis using next-generation sequencing. *Curr Opin Biotechnol*. 2013;24(1):22–30.

Funding

This work was funded in part by a grant from the Lung Cancer Research Foundation (to HK), Jimmy Lane Hewlett Fund for Lung Cancer Research (to IIW), NCI lung cancer SPORE P50 CA70907 (to IIW), Depart of Defense grants W81XWH-04-1-0142 (to WKH and IIW) and W81XWH-10-1-1007 (to HK, WKH, and IIW), and by the institutional Cancer Center Support grant CA16672.

Affiliations of authors: Department of Translational Molecular Pathology (HK, JF, TM, MMG, C-WC, ZC, GM, IIW), Department of Thoracic/Head and Neck Medical Oncology (MK, WKH, IIW), Department of Pathology (NK, CM), Department of Bioinformatics (S-YY, LS, JW, KRC), University of Texas MD Anderson Cancer Center, Houston, TX; Section of Computational Medicine, Department of Medicine, Boston University, Boston, MA (ACG, AS).